

JC07 Rec'd PCT/PTO 25 JAN 2002

U S DEPARTMENT OF COMMERCE PATENT AND TRADEMARK OFFICE FORM PTO-1390 (REV 12-2001)		ATTORNEY'S DOCKET NUMBER WFU.99-35
TRANSMITTAL LETTER TO THE UNITED STATES DESIGNATED/ELECTED OFFICE (DO/EO/US) CONCERNING A FILING UNDER 35 U.S.C. 371		
INTERNATIONAL APPLICATION NO. PCT/US00/20400	INTERNATIONAL FILING DATE 27 July 2000	PRIORITY DATE CLAIMED 27 July 1999
TITLE OF INVENTION METHODS AND COMPOSITIONS FOR DIAGNOSING PALMOPLANTAR KERATODERMAS AND DYSPLASIAS AND OTHER PERIODONTAL DISEASES		
APPLICANT(S) FOR DO/EO/US HART, Thomas C.		
Applicant herewith submits to the United States Designated/Elected Office (DO/EO/US) the following items and other information:		
<p>1. <input checked="" type="checkbox"/> This is a FIRST submission of items concerning a filing under 35 U.S.C. 371.</p> <p>2. <input type="checkbox"/> This is a SECOND or SUBSEQUENT submission of items concerning a filing under 35 U.S.C. 371.</p> <p>3. <input checked="" type="checkbox"/> This is an express request to begin national examination procedures (35 U.S.C. 371(f)). The submission must include items (5), (6), (9) and (21) indicated below.</p> <p>4. <input checked="" type="checkbox"/> The US has been elected by the expiration of 19 months from the priority date (Article 31).</p> <p>5. <input checked="" type="checkbox"/> A copy of the International Application as filed (35 U.S.C. 371(c)(2))</p> <ul style="list-style-type: none"> a. <input type="checkbox"/> is attached hereto (required only if not communicated by the International Bureau). b. <input checked="" type="checkbox"/> has been communicated by the International Bureau. c. <input type="checkbox"/> is not required, as the application was filed in the United States Receiving Office (RO/US) <p>6. <input type="checkbox"/> An English language translation of the International Application as filed (35 U.S.C. 371(c)(2)).</p> <ul style="list-style-type: none"> a. <input type="checkbox"/> is attached hereto. b. <input type="checkbox"/> has been previously submitted under 35 U.S.C. 154(d)(4). <p>7. <input checked="" type="checkbox"/> Amendments to the claims of the International Application under PCT Article 19 (35 U.S.C. 371(c)(3))</p> <ul style="list-style-type: none"> a. <input checked="" type="checkbox"/> are attached hereto (required only if not communicated by the International Bureau). b. <input type="checkbox"/> have been communicated by the International Bureau. c. <input type="checkbox"/> have not been made; however, the time limit for making such amendments has NOT expired. d. <input type="checkbox"/> have not been made and will not be made. <p>8. <input type="checkbox"/> An English language translation of the amendments to the claims under PCT Article 19 (35 U.S.C. 371(c)(3)).</p> <p>9. <input checked="" type="checkbox"/> An oath or declaration of the inventor(s) (35 U.S.C. 371(c)(4)).</p> <p>10. <input type="checkbox"/> An English language translation of the annexes of the International Preliminary Examination Report under PCT Article 36 (35 U.S.C. 371(c)(5)).</p>		
<p>Items 11 to 20 below concern document(s) or information included:</p> <p>11. <input type="checkbox"/> An Information Disclosure Statement under 37 CFR 1.97 and 1.98.</p> <p>12. <input type="checkbox"/> An assignment document for recording. A separate cover sheet in compliance with 37 CFR 3.28 and 3.31 is included.</p> <p>13. <input checked="" type="checkbox"/> A FIRST preliminary amendment.</p> <p>14. <input type="checkbox"/> A SECOND or SUBSEQUENT preliminary amendment.</p> <p>15. <input type="checkbox"/> A substitute specification.</p> <p>16. <input type="checkbox"/> A change of power of attorney and/or address letter.</p> <p>17. <input type="checkbox"/> A computer-readable form of the sequence listing in accordance with PCT Rule 13ter.2 and 35 U.S.C. 1.821 - 1.825.</p> <p>18. <input type="checkbox"/> A second copy of the published international application under 35 U.S.C. 154(d)(4).</p> <p>19. <input type="checkbox"/> A second copy of the English language translation of the international application under 35 U.S.C. 154(d)(4).</p> <p>20. <input checked="" type="checkbox"/> Other items or information: Copy of Form PCT/IB/308</p>		

U.S. APPLICATION NO. (if known, see 37 CFR 1.5)	INTERNATIONAL APPLICATION NO PCT/US00/20400	ATTORNEY'S DOCKET NUMBER WFU 99-35																									
<p><input checked="" type="checkbox"/> The following fees are submitted.</p> <p>BASIC NATIONAL FEE (37 CFR 1.492 (a) (1) - (5)):</p> <p>Neither international preliminary examination fee (37 CFR 1.482) nor international search fee (37 CFR 1.445(a)(2)) paid to USPTO and International Search Report not prepared by the EPO or JPO \$1040.00</p> <p>International preliminary examination fee (37 CFR 1.482) not paid to USPTO but International Search Report prepared by the EPO or JPO \$890.00</p> <p>International preliminary examination fee (37 CFR 1.482) not paid to USPTO but international search fee (37 CFR 1.445(a)(2)) paid to USPTO \$740.00</p> <p>International preliminary examination fee (37 CFR 1.482) paid to USPTO but all claims did not satisfy provisions of PCT Article 33(1)-(4) \$710.00</p> <p>International preliminary examination fee (37 CFR 1.482) paid to USPTO and all claims satisfied provisions of PCT Article 33(1)-(4) \$100.00</p> <p>ENTER APPROPRIATE BASIC FEE AMOUNT =</p>		CALCULATIONS PTO USE ONLY																									
<p>Surcharge of \$130.00 for furnishing the oath or declaration later than <input type="checkbox"/> 20 <input type="checkbox"/> 30 months from the earliest claimed priority date (37 CFR 1.492(e))</p>		\$ 100.00																									
<table border="1"> <tr> <th>CLAIMS</th> <th>NUMBER FILED</th> <th>NUMBER EXTRA</th> <th>RATE</th> <th>\$</th> </tr> <tr> <td>Total claims</td> <td>40 - 20 =</td> <td>20</td> <td>x \$18.00</td> <td>\$ 360.00</td> </tr> <tr> <td>Independent claims</td> <td>7 - 3 =</td> <td>4</td> <td>x \$84.00</td> <td>\$ 336.00</td> </tr> <tr> <td colspan="2">MULTIPLE DEPENDENT CLAIM(S) (if applicable)</td> <td></td> <td>+ \$280.00</td> <td>\$ 0</td> </tr> <tr> <td colspan="4">TOTAL OF ABOVE CALCULATIONS =</td> <td>\$ 796.00</td> </tr> </table>		CLAIMS	NUMBER FILED	NUMBER EXTRA	RATE	\$	Total claims	40 - 20 =	20	x \$18.00	\$ 360.00	Independent claims	7 - 3 =	4	x \$84.00	\$ 336.00	MULTIPLE DEPENDENT CLAIM(S) (if applicable)			+ \$280.00	\$ 0	TOTAL OF ABOVE CALCULATIONS =				\$ 796.00	
CLAIMS	NUMBER FILED	NUMBER EXTRA	RATE	\$																							
Total claims	40 - 20 =	20	x \$18.00	\$ 360.00																							
Independent claims	7 - 3 =	4	x \$84.00	\$ 336.00																							
MULTIPLE DEPENDENT CLAIM(S) (if applicable)			+ \$280.00	\$ 0																							
TOTAL OF ABOVE CALCULATIONS =				\$ 796.00																							
<p><input checked="" type="checkbox"/> Applicant claims small entity status See 37 CFR 1.27. The fees indicated above are reduced by 1/2.</p>		\$ 398.00																									
		SUBTOTAL =	\$ 398.00																								
<p>Processing fee of \$130.00 for furnishing the English translation later than <input type="checkbox"/> 20 <input type="checkbox"/> 30 months from the earliest claimed priority date (37 CFR 1.492(f)).</p>		\$ 0																									
		TOTAL NATIONAL FEE =	\$ 398.00																								
<p>Fee for recording the enclosed assignment (37 CFR 1.21(h)). The assignment must be accompanied by an appropriate cover sheet (37 CFR 3.28, 3.31). \$40.00 per property +</p>		\$ 0																									
		TOTAL FEES ENCLOSED =	\$ 398.00																								
		Amount to be refunded:	\$																								
		charged:	\$																								
<p>a. <input checked="" type="checkbox"/> A check in the amount of \$ <u>398.00</u> to cover the above fees is enclosed.</p> <p>b. <input type="checkbox"/> Please charge my Deposit Account No. _____ in the amount of \$ _____ to cover the above fees. A duplicate copy of this sheet is enclosed.</p> <p>c. <input checked="" type="checkbox"/> The Commissioner is hereby authorized to charge any additional fees which may be required, or credit any overpayment to Deposit Account No. <u>04-1406</u>. A duplicate copy of this sheet is enclosed.</p> <p>d. <input type="checkbox"/> Fees are to be charged to a credit card. WARNING: Information on this form may become public. Credit card information should not be included on this form. Provide credit card information and authorization on PTO-2038.</p>																											
<p>NOTE: Where an appropriate time limit under 37 CFR 1.494 or 1.495 has not been met, a petition to revive (37 CFR 1.137 (a) or (b)) must be filed and granted to restore the application to pending status.</p>																											
<p>SEND ALL CORRESPONDENCE TO</p> <p>Kathleen D. Rigaut, Ph.D., J.D. DANN, DORFMAN, HERRELL AND SKILLMAN 1601 Market Street, Suite 720 Philadelphia, Pennsylvania 19103 U.S.A.</p>																											
 <p>SIGNATURE Kathleen D. Rigaut, Ph.D., J.D. NAME 43,047 REGISTRATION NUMBER</p>																											

JC14 Rec'd PCT/PTO 28 MAY 2002

FES

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re the Application of)
Thomas C. Hart)
Serial No. 10/031,889)
Filed: January 25, 2002)
For: "Methods and Compositions)
For Diagnosing)
Palmoplantar Keratodermas)
And Dysplasias and Other)
Periodontal Diseases")

SUBMISSION OF SEQUENCE LISTING UNDER 37 C.F.R. §§1.821-1.825

AND PRELIMINARY AMENDMENT

This submission is in response to the April 22, 2002 Office Communication enclosing a Notification Of Missing Requirements Under 35 U.S.C. 371 In The United States Designated/Elected Office.

To comply with the requirements under 37 C.F.R. §§1.821-1.825, submitted herewith is a sequence listing of the nucleotide and amino acid sequences presented in the above-referenced application. The sequence listing is being submitted in both paper copy and computer-readable form. Applicants respectfully request entry of the sequence listing into the above identified patent application. The undersigned hereby verifies that the paper copy and computer readable form of the sequence listing are identical and do not contain any new matter.

In the event that a fee is required, the Commissioner is authorized to charge the account of the undersigned, Account No. 04-1406. A duplicate copy of this sheet is enclosed.

Respectfully submitted,

DANN, DORFMAN, HERRELL AND SKILLMAN
A Professional Corporation

By


Kathleen D. Rigaut, Ph.D., J.D.
PTO Registration No. 43,047

Telephone: (215) 563-4100

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re the Application of)
Thomas C. Hart et al.)
Serial No. Not yet assigned)
Filed: January 25, 2002)
For: "Methods and Compositions)
For Diagnosing)
Palmoplantar Keratodermas)
And Dysplasias and Other)
Periodontal Diseases")

The present application is based on International Application PCT/US00/20400. Before calculation of the filing fee, please amend the above-referenced patent application as follows:

In the specification:

At page 1, line 2, please insert the following priority claim:

-- This application is a 35 U.S.C. §371 application which claims priority to PCT/US00/20400 filed July 27, 2000 which in turn claims priority to U.S. Provisional Applications, 60/145,644 and 60/165,016 filed July 27, 1999 and November 12, 1999 respectively, the disclosure of each of these applications being incorporated herein by reference.

Please amend the specification at page 80 to include a copy of the abstract which is attached hereto on a separate sheet.

In the claims:

25. The method of claim 13 wherein a germline alteration is detected by obtaining a first CTSC gene fragment from (a) CTSC gene genomic DNA isolated from said sample, (b) CTSC RNA

isolated from said sample or (c) CTSC cDNA made from mRNA isolated from said sample and a second CTSC gene fragment from a CTSC allele specific for one of said alterations, said second fragment corresponding to said first fragment, forming single-stranded DNA from said first CTSC gene fragment and from said second CTSC gene fragment, forming a heteroduplex consisting of single-stranded DNA from said first CTSC gene fragment and single-stranded DNA from said second CTSC gene fragment and analyzing for the presence of a mismatch in said heteroduplex, wherein no mismatch indicates the presence of said alteration.

26. A method as claimed in claim 13, wherein said germline alteration comprises a substitution of a C for a T at nucleotide position 856 in Exon 6, thereby replacing a codon encoding glutamine for a stop codon.

27. A method as claimed in claim 13, wherein said germline alteration comprises a substitution of an A for a G at nucleotide position 857 in Exon 6, thereby replacing a codon encoding glutamine for an arginine encoding codon.

28. A method as claimed in claim 13, wherein said germline alteration comprises a deletion of an A at nucleotide position 1047 in Exon 7, thereby causing a frameshift and a premature stop codon.

29. A method as claimed in claim 13, wherein said germline alteration comprises a deletion of a dinucleotide CT at nucleotide positions 1028 and 1029 in Exon 7, thereby causing a premature stop codon.

30. A method as claimed in claim 13, wherein said germline alteration comprises a substitution of a G for a A at nucleotide position 1286 in Exon 7, thereby replacing a

tryptophan codon with a premature stop codon.

31. A method as claimed in claim 13, wherein said germline alteration comprises a substitution of a C for a T at nucleotide position 1015 in Exon 7, thereby replacing a codon encoding arginine for a cysteine encoding codon.

32. A method as claimed in claim 13, wherein said germline alteration comprises a substitution of an A for a G at nucleotide position 1019 in Exon 7, thereby replacing a codon encoding tyrosine for a cysteine encoding codon.

33. A method as claimed in claim 13, wherein said germline alteration comprises a substitution of an A for a G at nucleotide position 1040 in Exon 7, thereby replacing a codon encoding tyrosine for a cysteine encoding codon.

34. A method for detecting a germline alteration in a CTSC human encoding nucleic acid, said method comprising comparing a sequence of a CTSC DNA or CTSC RNA from a human sample with an isolated wild type CTSC sequence as provided in SEQ ID NO:1.

35. A method as claimed in claim 34, wherein stability of said altered CTSC mRNA is compared with stability of wild type CTSC mRNA.

36. A method as claimed in claim 34, further comprising expressing an altered CTSC protein from said altered CTSC encoding nucleic acid and comparing cathepsin C enzymatic activity of said altered CTSC protein to enzymatic activity of wild-type cathepsin C.

37. A kit for detecting the presence of an altered CTSC encoding nucleic acid in a biological sample, comprising:

- i) oligonucleotides which specifically hybridize with CTSC encoding nucleic acids having the alterations set forth in Table 1;
- ii) reaction buffer; and
- iii) an instruction sheet.

38. A kit as claimed in claim 37, wherein said oligonucleotide contains a tag.

39. A kit for detecting the presence an altered CTSC encoding nucleic acid in a biological sample, comprising:

- i) antibodies immunologically specific for the altered CTSC proteins of the invention;
- ii) a solid support with immobilized CTSC antigens as a positive control; and
- iii) an instruction sheet.

40. A kit as claimed in claim 39, wherein said antibody contains a tag.

REMARKS

The purpose of this preliminary amendment is to 1) correct a minor error in the numbering of the claims; 2) insert a priority claim into the specification and 3) provide a copy of the abstract on a separate sheet.

Favorable consideration leading to prompt allowance of the present application is respectfully requested.

Respectfully submitted,
DANN, DORFMAN, HERRELL AND SKILLMAN
A Professional Corporation

By 
Kathleen D. Rigant

Kathleen D. Rigaut, Ph.D., J.D.
PTO Registration No. 43,047

Telephone: (215) 563-4100

Enclosures: Marked up copy of the claims showing where
amendments have been made

Copy of an abstract of the specification on a
separate sheet

Marked up draft of claims:

[26] 25. The method of claim 13 wherein a germline alteration is detected by obtaining a first CTSC gene fragment from (a) CTSC gene genomic DNA isolated from said sample, (b) CTSC RNA isolated from said sample or (c) CTSC cDNA made from mRNA isolated from said sample and a second CTSC gene fragment from a CTSC allele specific for one of said alterations, said second fragment corresponding to said first fragment, forming single-stranded DNA from said first CTSC gene fragment and from said second CTSC gene fragment, forming a heteroduplex consisting of single-stranded DNA from said first CTSC gene fragment and single-stranded DNA from said second CTSC gene fragment and analyzing for the presence of a mismatch in said heteroduplex, wherein no mismatch indicates the presence of said alteration.

[27] 26. A method as claimed in claim 13, wherein said germline alteration comprises a substitution of a C for a T at nucleotide position 856 in Exon 6, thereby replacing a codon encoding glutamine for a stop codon.

[28] 27. A method as claimed in claim 13, wherein said germline alteration comprises a substitution of an A for a G at nucleotide position 857 in Exon 6, thereby replacing a codon encoding glutamine for an arginine encoding codon.

[29] 28. A method as claimed in claim 13, wherein said germline alteration comprises a deletion of an A at nucleotide position 1047 in Exon 7, thereby causing a frameshift and a premature stop codon.

[30] 29. A method as claimed in claim 13, wherein said germline alteration comprises a deletion of a dinucleotide CT at nucleotide positions 1028 and 1029 in Exon 7, thereby

causing a premature stop codon.

[31] 30. A method as claimed in claim 13, wherein said germline alteration comprises a substitution of a G for a A at nucleotide position 1286 in Exon 7, thereby replacing a tryptophan codon with a premature stop codon.

[32] 31. A method as claimed in claim 13, wherein said germline alteration comprises a substitution of a C for a T at nucleotide position 1015 in Exon 7, thereby replacing a codon encoding arginine for a cysteine encoding codon.

[33] 32. A method as claimed in claim 13, wherein said germline alteration comprises a substitution of an A for a G at nucleotide position 1019 in Exon 7, thereby replacing a codon encoding tyrosine for a cysteine encoding codon.

[34] 33. A method as claimed in claim 13, wherein said germline alteration comprises a substitution of an A for a G at nucleotide position 1040 in Exon 7, thereby replacing a codon encoding tyrosine for a cysteine encoding codon.

[35] 34. A method for detecting a germline alteration in a CTSC human encoding nucleic acid, said method comprising comparing a sequence of a CTSC DNA or CTSC RNA from a human sample with an isolated wild type CTSC sequence as provided in SEQ ID NO:1.

[36] 35. A method as claimed in claim 3[5]4, wherein stability of said altered CTSC mRNA is compared with stability of wild type CTSC mRNA.

[37] 36. A method as claimed in claim 3[5]4, further comprising expressing an altered CTSC protein from said altered CTSC encoding nucleic acid and comparing cathepsin C

enzymatic activity of said altered CTSC protein to enzymatic activity of wild-type cathepsin C.

[38] 37. A kit for detecting the presence of an altered CTSC encoding nucleic acid in a biological sample, comprising:

- i) oligonucleotides which specifically hybridize with CTSC encoding nucleic acids having the alterations set forth in Table 1;
- ii) reaction buffer; and
- iii) an instruction sheet.

[39] 38. A kit as claimed in claim 3[8]7, wherein said oligonucleotide contains a tag.

[40] 39. A kit for detecting the presence an altered CTSC encoding nucleic acid in a biological sample, comprising:

- i) antibodies immunologically specific for the altered CTSC proteins of the invention;
- ii) a solid support with immobilized CTSC antigens as a positive control; and
- iii) an instruction sheet.

[41] 40. A kit as claimed in claim [40]39, wherein said antibody contains a tag.

ABSTRACT

Compositions relating to mutations in the CTSC gene associated with palmoplantar keradermas are disclosed. Also provided are methods for using the compositions of the invention in diagnostic and prognostic screening assays.

WO 01/07663

PCT/US00/20400

METHODS AND COMPOSITIONS FOR DIAGNOSING PALMOPLANTAR
KERATODERMAS AND DYSPLASIAS AND OTHER PERIODONTAL
DISEASES

5

Pursuant to 35 U.S.C. §202(c), it is acknowledged that the U.S. Government has certain rights in the invention described herein, which was made in part with funds from NIDCR, Grant Numbers: DE11601 and DE12920.

FIELD OF THE INVENTION

The present invention relates to the fields of genetic screening and molecular biology. More specifically, the invention provides compositions and methods that may be used to advantage to isolate and detect a palmoplantar keratoderma predisposing gene, cathepsin C (CTSC), some mutant alleles of which cause susceptibility to certain pathological disorders, in particular Papillon-LeFevre Syndrome, Haim-Munk Syndrome and certain forms of early onset periodontal diseases. More specifically, the invention relates to germline mutations and functional polymorphisms in the CTSC gene and their use in the diagnosis of predisposition to palmoplantar ectodermal disorders/ dysplasias and periodontal diseases. The invention also relates to the therapy of palmoplantar ectodermal disorders/dysplasias and periodontal diseases which have a mutation or functional polymorphisms in the CTSC gene, including gene therapy, protein replacement therapy and protein mimetics. The invention further relates to the screening of drugs for treating and alleviating disease symptoms. Finally, the invention relates to the screening of the CTSC gene for disease-related mutations, which are useful for diagnosing the predisposition to additional disorders and dysplasias, including but not limited to prepubertal periodontitis, early onset periodontal disease or other forms of gum disease.

BACKGROUND OF THE INVENTION

Various publications or patents may be referenced in this application by numerals in parentheses to describe the state of the art to which the invention pertains. Full citations for these references are provided at the end of the specification. Each of these publications or patents is incorporated by reference herein.

Most forms of inflammatory periodontal disease can be successfully treated and managed. As a result, the ultimate goal of periodontal therapy has changed from that of simply arresting disease progression to one aimed at regenerating the supporting tissues. Unfortunately, not all forms of periodontal disease respond to treatment. Severe periodontitis that is resistant to conventional periodontal treatment has been recognized in a number of monogenic conditions. Certainly some of the most intriguing and dentally challenging of these conditions include Papillon-Lefevre syndrome (PLS), Haim-Munk syndrome (HMS) and periodontal diseases.

In 1924, Papillon and Lefevre described two siblings, the products of a first cousin mating, with a condition characterized by diffuse transgradient palmoplantar keratosis (PPK) and the premature loss of both the deciduous and permanent dentitions. This condition came to be known as Papillon-Lefevre syndrome and subsequently over 200 cases have been described. The hallmarks of PLS are palmoplantar keratosis and rapid periodontal destruction of both dentitions. An increased susceptibility to infection has been reported in approximately 20% of PLS patients. Additional findings include intracranial calcifications, retardation of the somatic development, follicular hyperkeratosis and onychogryphosis. Clinical findings reported in PLS patients suggest that the clinical expression of the condition is

highly variable. Unfortunately, to date, no pathognomonic disease marker exists allowing definitive diagnosis of PLS.

In 1965, Haim and Munk described an unusual syndrome in four siblings of a Jewish religious isolate from Cochin, India [21]. In addition to congenital palmoplantar keratosis and progressive early onset periodontal destruction, other clinical findings shared by these individuals included recurrent pyogenic skin infections, acroosteolysis, atrophic changes of the nails, arachnodactyly, and a peculiar radiographic deformity of the fingers consisting of tapered pointed phalangeal ends and a clawlike volar curve. Subsequently pes planus was reported to be associated with the syndrome [24]. This was the first reported association of these clinical findings, and the condition became known as Haim Munk syndrome, or keratosis palmoplantaris with periodontopathia and onychogryposis (HMS; MIM245010) [22]. Although the palmoplantar findings and severe periodontitis were suggestive of the Papillon-Lefevre syndrome (PLS; MIM245000) [3], the association of other clinical features, particularly nail deformities and arachnodactyly, argued that HMS was a distinct disorder.

PLS and HMS are classified as type IV palmoplantar ectodermal keratodermas [2]. The unique presence of severe, early onset periodontitis distinguishes PLS and HMS from other PPKs and raises the question of whether they result from the variable clinical expression of a common gene mutation, are allelic mutations at the same genetic locus, or result from expression of gene mutations at separate loci. Although Haim and Munk's initial report proposed HMS was a distinct entity, Hacham-Zadeh and co-workers referred to the disorder as Papillon-Lefevre syndrome and cited Gorlin's suggestion that HMS was a clinical variant of PLS [23,25]. In his review of PLS cases reported in the literature, Haenke

[5] summarizes an extensive list of clinical findings reported in PLS affected individuals, including increased susceptibility to infections, ectopic cranial calcifications and nail anomalies [5,26]. It is unclear if these additional clinical features are coincidental findings that may be segregating in a particular family or if they are etiologically related to a syndrome with a very variable clinical expression. Because PLS is an uncommon condition, and generally occurs only in a single generation it is difficult to determine if these occasional reports of associated clinical findings are etiologically related to PLS. Additionally, consanguinity is common among parents of PLS cases and therefore, it may be expected that an increased number of rare recessive conditions may be seen. Such is likely the case for the reports of mental retardation associated with PLS [5].

Pre-pubertal periodontitis (PPP) is a rare and rapidly progressive disease that results in destruction of the periodontal support of the primary dentition. The condition may be localized (usually to deciduous molars) or generalized. The localized form begins at approximately 4 years of age and is associated with only mild gingival inflammation in the presence of relatively little plaque. The generalized form begins earlier, immediately after eruption of the deciduous teeth. It is associated with severe gingival inflammation and hyperplasia, although significant gingival recession has also been described as an associated clinical feature. The attachment loss appears to be continuous rather than intermittent as with most other forms of periodontitis.

A varied clinical phenotype has been reported for PPP, probably reflecting the fact that the term PP describes an etiologically heterogeneous group of conditions that share an overlapping clinical presentation. Although PPP can occur as an isolated finding, many reports of PPP describe an increased

systemic susceptibility to infections. Children with PPP have frequent inner ear infections and infections of the respiratory tract [39,40]. Prepubertal periodontitis is known to be associated with Papillon Lefevre syndrome and with a number of systemic disease states that share an increased susceptibility to microbial infections.

To date, no pathognomonic disease marker exists for most PPKs allowing for definitive diagnosis. The present invention provides such a disease marker and methods of use thereof having diagnostic and prognostic utilities for several PPKs and many periodontal diseases.

SUMMARY OF THE INVENTION

The present invention provides compositions and methods which allow for genetic screening and diagnosis of certain palmoplantar keratodermas and periodontal disease states in affected individuals. In accordance with the present invention, it has been discovered that mutations or functional polymorphisms in the cathepsin C gene (CTSC) give rise to certain pathological conditions including PLS, HMS and periodontal diseases. Mutations or functional polymorphisms associated with the disease state are those which give rise to a altered, truncated, misfolded or otherwise non-functional CTSC polypeptides. Polymorphisms in the CTSC sequence which do not affect the nature of the encoded protein are not associated with PLS, HMS or periodontal disease.

Thus, in one embodiment of the invention, a method is provided for determining the presence of alterations in CTSC encoding nucleic acids which give rise to altered CTSC proteins. The wild-type CTSC nucleic acid sequence and its encoded amino acid sequence are known. See SEQ ID NOS: 1-3 provided herein. This sequence information facilitates the identification of genetic changes that give rise to aberrant CTSC proteins.

CTSC mutations specifically associated with PLS,

HMS and PPP are described herein and are set forth in Table 1. Accordingly in one embodiment of the invention, nucleic acid molecules encoding altered CTSC proteins are considered to be within the scope of the present invention. In a preferred embodiment of the invention, the altered CTSC nucleic acid has at least one of the alterations set forth in Table 1.

In a further embodiment of the invention, nucleic acid probes which specifically hybridize to the human altered CTSC-encoding nucleic acids described herein and not to wild-type CTSC encoding nucleic acids are provided. In a preferred embodiment, the probes specifically hybridize with altered CTSC encoding nucleic acids having at least one of the alterations set forth in Table 1.

In yet another embodiment of the invention, a mutated CTSC protein encoded by the altered CTSC encoding nucleic acids of the invention is provided. Preferably such CTSC proteins are encoded by a nucleic acid containing a mutation as set forth in Table 1. Also provided are assays for biochemically assessing altered cathepsin C activity. Antibodies immunologically specific for altered CTSC proteins are also contemplated to be within the scope of the present invention.

In another aspect of the invention, a method for detecting a germline alteration in a CTSC gene is provided. In a preferred embodiment the alteration is selected from the group consisting of the alterations set forth in Table 1. The method comprises analyzing a sequence of a CTSC gene or CTSC RNA from a human sample or analyzing a sequence of CTSC cDNA made from mRNA from a human sample and comparing sequences so isolated to the wild type sequence encoding CTSC. Inasmuch as certain alterations of the CTSC coding sequence may not alter the function of CTSC, methods are provided for assessing the enzymatic activity of proteins encoded by nucleic acid molecules which do not possess the wild

type CTSC sequence.

In yet another embodiment of the invention, kits are provided for detecting the presence of an altered CTSC encoding nucleic acids in a biological sample. An exemplary kit comprises the following: i) oligonucleotides which specifically hybridize with CTSC encoding nucleic acids having the alterations set forth in Table 1; ii) reaction buffer; and iii) an instruction sheet. Kits for detecting the presence of an altered CTSC proteins in a biological sample are also provided. Exemplary kits for this purpose comprise: i) antibodies immunologically specific for the altered CTSC proteins of the invention; ii) a solid support with immobilized CTSC antigens as a positive control; and iii) an instruction sheet. Optionally, anti-CTSC antibodies used for this purpose may contain a detectable label or tag for use in isolating or detecting immune complexes.

Various terms relating to the biological molecules and cells of the present invention are used throughout the specifications and claims.

With reference to nucleic acids used in the invention, the term "isolated nucleic acid" is sometimes employed. This term, when applied to DNA, refers to a DNA molecule that is separated from sequences with which it is immediately contiguous (in the 5' and 3' directions) in the naturally occurring genome of the organism from which it was derived. For example, the "isolated nucleic acid" may comprise a DNA molecule inserted into a vector, such as a plasmid or virus vector, or integrated into the genomic DNA of a prokaryote or eucaryote. An "isolated nucleic acid molecule" may also comprise a cDNA molecule. An isolated nucleic acid molecule inserted into a vector is also sometimes referred to herein as a *Arecombinant*® nucleic acid molecule.

With respect to RNA molecules, the term "isolated

nucleic acid" primarily refers to an RNA molecule encoded by an isolated DNA molecule as defined above. Alternatively, the term may refer to an RNA molecule that has been sufficiently separated from RNA molecules with which it would be associated in its natural state (i.e., in cells or tissues), such that it exists in a "substantially pure" form (the term "substantially pure" is defined below).

With respect to protein, the term "isolated protein" or "isolated and purified protein" is sometimes used herein. This term refers primarily to a protein produced by expression of an isolated nucleic acid molecule of the invention. Alternatively, this term may refer to a protein which has been sufficiently separated from other proteins with which it would naturally be associated, so as to exist in "substantially pure" form.

The term "substantially pure" refers to a preparation comprising at least 50-60% by weight the compound of interest (e.g., nucleic acid, oligonucleotide, protein, etc.). More preferably, the preparation comprises at least 75% by weight, and most preferably 90-99% by weight, the compound of interest. Purity is measured by methods appropriate for the compound of interest (e.g. chromatographic methods, agarose or polyacrylamide gel electrophoresis, HPLC analysis, and the like).

With respect to antibodies, the term "immunologically specific" refers to antibodies that bind to one or more epitopes of a protein of interest, but which do not substantially recognize and bind other molecules in a sample containing a mixed population of antigenic biological molecules.

With respect to single stranded nucleic acids, particularly oligonucleotides, the term "specifically hybridizing" refers to the association between two single-stranded nucleotide molecules of sufficiently complementary sequence to permit such hybridization

under pre-determined conditions generally used in the art (sometimes termed "substantially complementary"). In particular, the term refers to hybridization of an oligonucleotide with a substantially complementary sequence contained within a single-stranded DNA or RNA molecule of the invention, to the substantial exclusion of hybridization of the oligonucleotide with single-stranded nucleic acids of non-complementary sequence. Appropriate conditions enabling specific hybridization of single stranded nucleic acid molecules of varying complementarity are well known in the art.

For instance, one common formula for calculating the stringency conditions required to achieve hybridization between nucleic acid molecules of a specified sequence homology is set forth below (Sambrook et al., 1989):

$$T_m = 81.5^{\circ}\text{C} + 16.6\log [\text{Na}^+] + 0.41(\% \text{ G+C}) - 0.63 (\% \text{ formamide}) - 600/\# \text{bp in duplex}$$

As an illustration of the above formula, using $[\text{Na}^+] = [0.368]$ and 50% formamide, with GC content of 42% and an average probe size of 200 bases, the T_m is 57°C . The T_m of a DNA duplex decreases by $1 - 1.5^{\circ}\text{C}$ with every 1% decrease in homology. Thus, targets with greater than about 75% sequence identity would be observed using a hybridization temperature of 42°C .

The term "promoter region" refers to the transcriptional regulatory regions of a gene, which may be found at the 5' or 3' side of the coding region, or within the coding region, or within introns. In the present invention, the use of SV40, TK, Albumin, SP6, T7 gene promoters, among others, is contemplated. Specific promoters for the yeast and mammalian expression systems of the invention are available and known to those of ordinary skill in the art.

The term "operably linked" means that the
q

regulatory sequences necessary for expression of the coding sequence are placed in the DNA molecule in the appropriate positions relative to the coding sequence so as to enable expression of the coding sequence. This same definition is sometimes applied to the arrangement of transcription units and other transcription control elements (e.g. enhancers) in an expression vector.

The phrase "functional polymorphism" refers to a change in wild type CTSC coding sequence giving rise to altered cathepsin C activity as assayed using conventional methods.

The term "oligonucleotide," as used herein refers to primers and probes of the present invention, and is defined as a nucleic acid molecule comprised of two or more ribo- or deoxyribonucleotides, preferably more than three. The exact size of the oligonucleotide will depend on various factors and on the particular application and use of the oligonucleotide.

The term "probe" as used herein refers to an oligonucleotide, polynucleotide or nucleic acid, either RNA or DNA, whether occurring naturally as in a purified restriction enzyme digest or produced synthetically, which is capable of annealing with or specifically hybridizing to a nucleic acid with sequences complementary to the probe. A probe may be either single-stranded or double-stranded. The exact length of the probe will depend upon many factors, including temperature, source of probe and use of the method. For example, for diagnostic applications, depending on the complexity of the target sequence, the oligonucleotide probe typically contains 15-25 or more nucleotides, although it may contain fewer nucleotides. The probes herein are selected to be complementary to different strands of a particular target nucleic acid sequence. This means that the probes must be sufficiently complementary so as to be able to "specifically hybridize" or anneal with their respective target

strands under a set of pre-determined conditions. Therefore, the probe sequence need not reflect the exact complementary sequence of the target. For example, a non-complementary nucleotide fragment may be attached to the 5' or 3' end of the probe, with the remainder of the probe sequence being complementary to the target strand. Alternatively, non-complementary bases or longer sequences can be interspersed into the probe, provided that the probe sequence has sufficient complementarity with the sequence of the target nucleic acid to anneal therewith specifically.

The term "primer" as used herein refers to an oligonucleotide, either RNA or DNA, either single-stranded or double-stranded, either derived from a biological system, generated by restriction enzyme digestion, or produced synthetically which, when placed in the proper environment, is able to functionally act as an initiator of template-dependent nucleic acid synthesis. When presented with an appropriate nucleic acid template, suitable nucleoside triphosphate precursors of nucleic acids, a polymerase enzyme, suitable cofactors and conditions such as a suitable temperature and pH, the primer may be extended at its 3' terminus by the addition of nucleotides by the action of a polymerase or similar activity to yield an primer extension product. The primer may vary in length depending on the particular conditions and requirement of the application. For example, in diagnostic applications, the oligonucleotide primer is typically 15-25 or more nucleotides in length. The primer must be of sufficient complementarity to the desired template to prime the synthesis of the desired extension product, that is, to be able anneal with the desired template strand in a manner sufficient to provide the 3' hydroxyl moiety of the primer in appropriate juxtaposition for use in the initiation of synthesis by a polymerase or similar enzyme. It is not required that the primer

sequence represent an exact complement of the desired template. For example, a non-complementary nucleotide sequence may be attached to the 5' end of an otherwise complementary primer. Alternatively, non-complementary bases may be interspersed within the oligonucleotide primer sequence, provided that the primer sequence has sufficient complementarity with the sequence of the desired template strand to functionally provide a template-primer complex for the synthesis of the extension product.

BRIEF DESCRIPTION OF THE DRAWINGS

Figures 1A-1E are a series of clinical photographs showing palmoplantar keratosis and periodontal disease in PLS study patient. Fig. 1A: palmar hyperkeratotic lesions; Fig. 1B: plantar hyperkeratotic lesions; Fig. 1C: hyperkeratotic lesions affecting the knees; Fig. 1D: periodontitis involving erupting permanent dentition and Fig. 1E: periapical radiographs showing severe alveolar bone loss affecting erupting permanent teeth.

Figure 2 shows haplotype data for chromosome 11q short tandem repeat polymorphisms (STRP) markers spanning the PLS gene locus. Segments which are likely to be homozygous by descent are boxed. Arrows indicate recombinant events. Individuals 7 and 22 share a common haplotype for D11S1979, D11S1887, D11S1780, D11S1367, D11S931, and D11S4175.

Figure 3 depicts pedigree and sequence analysis of CTSC exon 6. The numbering of the wildtype sequence shown above the figure is based upon the genomic sequence of CTSC. See SEQ ID NO: 1. Circles represent females and squares represent males. Filled symbols indicate affected individuals. Half-shading indicates carriers based upon DNA sequencing results. All affected individuals are homozygous for the specific CTSC

mutations. Arrows indicate the position of the mutation. This family has a nonsense mutation (856 C->T) at codon 286 resulting in a truncated protein of 286 amino acids.

5 Figures 4A - 4D show pedigrees and sequence analysis of *CTSC* exon 7 for 4 Families with PLS. Symbols are as described for Figure 3. Fig. 4A: Family with a single base pair deletion (1047delA) of *CTSC* resulting in a frameshift and premature termination.
10 Fig. 4B: Family with a 2bp deletion (1028-1029delCT) of *CTSC* resulting in a frameshift and premature termination. Fig. 4C. and Fig. 4D: pedigrees of families with a nonsense mutation (1286G->A) at codon 429 resulting in a truncated protein of 428 amino acids. The father in family C is deceased and no sample was
15 available for analysis.

Figures 5A-5C is a schematic diagram of the *CTSC* gene showing the locations of the mutations described herein. Panel 5A. Genomic structure of *CTSC* gene with introns shown as solid lines and exons depicted as boxes. The 5' and 3' untranslated regions are shown as filled boxes. Panel 5B. Coding region of *CTSC* gene. The amino acid numbers are shown at the end of each exon.
20 Mutations listed in Table 1 are shown according to their genomic locations with Missense, *; Nonsense, ♦;
25 Insertion, ▀; and Deletion, □. The splicing site mutation is indicated by an arrow. Panel 5C. Subunit structure of *CTSC* polypeptide with SP, signal peptide; P1, 13.5 kDa pro-region; P2, 10 kDa pro-region; H, heavy chain; and L, light chain. The 10 kDa pro-region is cleaved out upon activation. The disulfide bond within the 13.5 pro-region is shown. The glycosylation sites are indicated by filled circle and arrows indicate the active sites.
30
35

Figures 6A-6F show a series of micrographs

depicting the clinical and radiographic findings in Haim Munk syndrome. Fig. 6A: dermal involvement of fingers in individual #34. Fig. 6B: Individual #34 radiograph of terminal phalanges of the fingers showing marked thinning increasing towards the distal, tapering pointed ends showing a claw-like volar bend. Fig. 6C: Individual #17 palmar keratosis; Fig. 6D: Individual #17 plantar keratosis; Fig. 6E: Individual #17 gingival inflammation; Fig. 6F: Individual #17 radiograph showing alveolar bone destruction associated with gingival inflammation shown in 6D.

Figure 7A shows pedigree of Cochin descendants segregating Haim Munk syndrome (HMS). Numbered individuals have been analyzed for the current study. Circles = females, squares = males, shaded symbols = HMS affected individuals. Double lines = consanguinity. Individuals #10,11 * = second cousins. Numbers inside circles, squares and diamonds indicate the number of additional offspring not examined in this study. Sibships described in previous reports are indicated and referenced below the pedigrees. The subjects of Haim and Munks original report (1965) are individuals 33, 34, 35, and 36. Half-shading indicates carriers based upon DNA sequencing and/or restriction enzyme analysis. Unshaded numbered individuals represent non-carriers based upon DNA analysis. Figure 7B shows a pedigree of a Turkish family segregating PLS. Numbered individuals were available for study. Half-shading indicates carrier based upon DNA analysis. Individual 77 is a non-carrier based upon DNA sequencing.

Figures 8A and 8B show the results of sequence analysis of exon 6 of CTSC. The numbering of the wildtype sequence is based upon the cDNA sequence of CTSC. See SEQ ID NO: 1. Fig. 8A: Family A (Cochin isolate diagnosed with Haim Munk syndrome) from Figure

1. Affected individuals are homozygous for a 857A->G missense mutation which results in a conserved glutamine being changed to an arginine (Q286R). Representative sequences are shown for individuals #36 (affected) and #31(carrier). Fig. 8B. Family B from Figure 1.

5 Affected individuals are homozygous for a 856C->T nonsense mutation which results in a premature stop codon at position 286 (Q286X). The Q286X mutation has been previously reported in an inbred Turkish family

10 [12].

Figure 9 depicts a gel showing the results of restriction enzyme analysis of Q286R mutation in Family A of Fig. 7. A 465 bp fragment of exon 6 was amplified and subjected to restriction digestion with AvaI as described under methods. The Q286R mutation introduces a new AvaI site. After digestion and electrophoresis through 1.8% agarose gels, wildtype individuals exhibit bands of 465 bp, affected individuals have bands of 404 and 61 bp, and carriers have bands of 465, 404, and 61 bp. M. 1 kb ladder (Gibco). Lane 1. Individual #5 uncut, demonstrating 465bp amplicon. Lane 2. Individual #5 cut with AvaI. Only the 465 bp fragment is observed. Thus individual #5 has the wildtype sequence on both alleles. Lane 3. Individual #31 uncut. Lane 4. Individual # 31 cut with AvaI. The 465 and 404bp fragments are visible, confirming that individual #31 is a carrier of the Q286R, consistent with the sequencing results shown in Figure 3A. Lane 5. Individual #34 uncut. Lane 6. Individual #34 cut with AvaI. The 404 and 61bp fragments are indicated by arrows.

DETAILED DESCRIPTION OF THE INVENTION

The present invention relates generally to the field of human genetics. Specifically, the present invention relates to methods and materials used to isolate and detect mutated forms of the lysosomal

protease cathepsin C (CTSC) gene, associated with autosomal recessive disorders characterized by palmar hyperkeratosis and/or periodontitis. More specifically, the present invention relates to germline mutations in the CTSC gene and their use in the diagnosis of predisposition to such pathological conditions. Additionally, the invention relates to germline mutations in the CTSC gene in other palmoplantar ectodermal disorders and dysplasias and their use in the diagnosis and prognosis of such pathological conditions.

The invention also relates to the therapy of palmoplantar ectodermal disorders and dysplasias which have a mutation or functional polymorphism in the CTSC gene, including gene therapy, protein replacement therapy and protein mimetics. The invention further relates to the screening of drugs which may have therapeutic value. Biochemical assays are provided for the assessment of altered activity of aberrant CTSC enzymes encoded by the mutated CTSC encoding nucleic acids of the invention. Finally, the invention relates to the screening of the CTSC gene for mutations, which are useful for diagnosing the predisposition to ectodermal disorders and dysplasias.

The present invention provides an isolated polynucleotide comprising all, or a portion of the CTSC locus or of a mutated CTSC locus, preferably at least eight bases and not more than about 100 kb in length. Such polynucleotides may be antisense polynucleotides. The present invention also provides a recombinant construct comprising such an isolated polynucleotide, for example, a recombinant construct suitable for expression in a transformed host cell.

Also provided by the present invention are methods of detecting a polynucleotide comprising a portion of the CTSC locus or its expression product in an analyte. Such methods may further comprise the step of amplifying the portion of the CTSC locus, and may further include a

step of providing a set of polynucleotides which are primers for amplification of said portion of the CTSC locus. The method is useful for either diagnosis of the predisposition to PPKs or the diagnosis or prognosis of keratodermal disorders/dysplasias and periodontal diseases.

The present invention also provides isolated antibodies, preferably monoclonal antibodies, which specifically bind to an isolated polypeptide comprised of at least five amino acid residues encoded by the altered CTSC locus.

The present invention also provides kits for detecting in an analyte a polynucleotide comprising a portion of the CTSC locus, the kits comprising a polynucleotide complementary to the portion of the CTSC locus packaged in a suitable container, and instructions for its use.

The present invention further provides methods of preparing a polynucleotide comprising polymerizing nucleotides to yield a sequence comprised of at least eight consecutive nucleotides of the CTSC locus; and methods of preparing a polypeptide comprising polymerizing amino acids to yield a sequence comprising at least five amino acids encoded within the CTSC locus.

The present invention further provides methods of screening the CTSC gene to identify mutations. Such methods may further comprise the step of amplifying a portion of the CTSC locus, and may further include a step of providing a set of polynucleotides which are primers for amplification of said portion of the CTSC locus. Exemplary primers are set forth in Table A.

The method is useful for identifying mutations for use in either diagnosis of the predisposition to keratodermal disorders/dysplasias and periodontal diseases or the diagnosis of such disorders.

The present invention further provides methods of screening suspected CTSC mutant alleles and functional

polymorphisms to identify mutations in the CTSC gene. In addition, the present invention provides methods of screening drugs for therapy and to identify suitable drugs for restoring CTSC gene product function.

5 Finally, the present invention provides the means necessary for production of gene-based therapies directed at aberrant cells associated with keratodermal disorders and dysplasias. These therapeutic agents may take the form of polynucleotides comprising all or a portion of the CTSC locus placed in appropriate vectors or delivered to target cells in more direct ways such that the function of the CTSC protein is reconstituted. Therapeutic agents may also take the form of polypeptides based on either a portion of, or the entire 10 protein sequence of CTSC. These may functionally replace 15 the activity of CTSC in vivo.

It is a discovery of the present invention that mutations in the CTSC locus in the germline are indicative of a predisposition to keratodermal 20 disorders/dysplasias and periodontal diseases. The mutational events of the CTSC locus can involve deletions, insertions and point mutations within the coding sequence and the non-coding sequence.

A major gene locus associated with the keratodermal 25 disorders and dysplasias of the invention has been localized to a 2.8 cM interval on chromosome 11q14 of the human genome. This region contains a genetic locus, CTSC. The CTSC message is expressed at high levels in a variety of immune cells including polymorphonuclear 30 leukocytes, macrophages and their precursors. This gene is expressed in the palms, soles, knees, and oral keratinized gingiva.

The CTSC gene was originally reported to consist of 2 exons. US Provisional Application 60/165,016 from 35 which the present application claims priority, describes mutations in Exons 1 and 2. The mutations described actually fall within Exons 6 and 7. Reference numerals

to the altered amino acids are the same as those in US Provisional 60/165,016, only the nucleotide numbering has changed to reflect the actual genomic structure of the CTSC gene which is now known to contain 7 exons.
5 The sequence encoding the wild type human CTSC gene is provided below (SEQ ID NO: 1):

AGGGAGATAT AAGTGAATAA TTTGGACCTG CTCTCTTGA ATGTTTATAA TCTGGTGGAA
10 AAAAAATGGA CATATGAATA TTGATTGTG ACCAGTGCAA AGGGGGCAAA AATTCATATC
CCAAAGAAAA CGGGGACACA TCAGGTCTGT CTTGTTCATC ACTGTGTCCA CAGGCCCTGA
CACCTAGTAG GCTCAGTGGG AGAAAGGAGC CCCAATTACC AACAAAAGCC AGGAAAGAAC
GGGAGGCTCT TACGGAAAAG GGTGATACTT AAACTGAGCA AGGAGGCACC TGGAAATAGT
GCCACCTAAT AATTTTGCG GATCAGACTG ACACACTAGA ACGGTTCATC AGACCAGCCT
15 TCTCCCATTG GCTAGCTTCC TTCTCACCC TTCTCACCC GGGCAAGCCG CTTCCCTCT
CTGGGCCTCT TGCTTTCCCT CTGTAACATA AAAGGGTTG AGCAATATCA TCTCTGAGAG
CGCCATGTGT GTGCGTGCCTA GAGGGAAAAC CCCCACAACG CTAATACATC AAAACTGCAG
GTTTGCACAA AAACTGAATT CTGCTGAATG CAAACAGGCA AACAGCATTT ACCAGGAAAC
AAAACAAAAT CAAGCACATA AAAAAGTAGG AAGAGTTGGA AACACGGAAAGG AAGATAAGTT
20 CTCAAACAGC TGGAATAGTT GATGTTAGCT AGCGAAGTTT TTCAGAGGAA AAAACAAGAA
GTTGGTTATG AGGCAAGTGG ACCTGAGAAA AAAGACTAAA GGGGAAGAAT AGCAAGTAAA
ACAGAACTCC ACTTGCTAGA TCTCTCCCTC TGTCGCGCTC TTTCACCTGA CCCACTCCCT
TATTCCCCCC ACACCCCTTC CTTCTCTCCC TACGTTACCG CACAGGAACG AAGTCTGGGT
25 CATGTGCGGA CCGCTTGTGG CTCTTAAATC CTCTTTGT CACCCCTGGCC GTGCAAAATT
TTGAAACGTC CCTCGGCAAA AAAAATAAAA ATAAAAAAA AAAATCTGTC CCTGGCCTCT
TCCCTAGTTC TGGGTCCAGT TGCAGCCAAG TGAGGGGCAG CGCGCGCTCC CAAGTCCCCG
TTTCAGAGAC GCGCACGCGC CTGGCGCCCA ACCCCCCAATC CCCTGCTGCT CAGTGACCCC
GCCACGGGT TTCCGGGCCG GCGTAGCTAT TTCAAGGCGC GCGCTCGTG GTGGACTCAC
CGCTAGCCCG CAGCGCTCGG CTTCTGGTA ATTCTTCACC TCTTTCTCA GCTCCCTGCA
30 GCATGGGTGC TGGGCCCTCC TTGCTGCTCG CCGCCCTCCT GCTGCTTCTC TCCGGCGACG
GCGCCGTGCG CTGGCACACA CCTGCCAACT GCACCTATCT TGACCTGCTG GGCACCTGGG
TCTTCCAGGT GGGCTCCAGC GGTTCCCAGC GCGATGTCAA CTGCTCGGTT ATGGGTAAGC
CGCCGGCTCG GCAGTCCTCC GGTCGTCCT TTCTGCCCTT GAGCCCCCAA CGCAGCGCCA
CGCCAACCTAC CGCTTCCCCC CAGGCAGACG CTTGTGGTG GCCAGAGCAT CTTGACTGGA
35 TTCGGGGACC CTTGGGGACC TTCTTCCCCC CCAGGCTCGC GAAGTTAAAG TTCATCTGCT
GAGAACTTCT AACTCCACAC TTTCTTGGTT ATCTTGGGA CTCAACACTT TGATCAAGAA
CTTTTTATT CCTCCCGCTT AATTTTGTGTT GCTTTGAGAG AGACTTGGGA ACTGCAATCG
TTTGGTTCTC CAGTCCGATC TGGTAGCGTT ATTTTAAAA TTTATTTTA TTTTTTATT
CTATTTTACT AGTGAAGATA GATGAGCTCA GAGACTCTCG AGGATATAGC ATGAAGTTTT
40 CTCTTTTGT TAGATGGTGG GAAAGGGACT TTCTGCCAG CGATTTGGT TTGAGCGGGT
GTTGATGAGT ACTAGAAAAC GGCTAGTACC ACTCTGCATT GTTTCATGCA TTGCAAGGAG
GTAAAAAATT TTTAAAAAAT TAATAACAAA GAAAACCTAA CTCTGAACCT AGTAATTAGA
AATGCCAGA GTCTGCACAA TGTTGGCTC ATGGAAGGCT CTCATAAAAT ACCTAGTGT
TGAACATACT GGAGATATTC CATATGCCTT CAGATAACAT GGTTACCCCT AGAACAAAGA

ACCTAGTGAA GGGGGTGGGG TGGAAAAAAG ACAATCACTA GTTAGAAGAG TCACACTGTG
GTCTACCCAA ACCCTCTTAC AGCCTGTGAC TTCTGCAGAG TCAGTAAAAA ATCAGCTATA
ATTTTCTTGT CAACAGAAC AAGATTAAGT CATTGTTAC TAAAGAAATA CCTTTTTAAC
5 TGACTGTATA GCATTCATA ATCCTGAACA TTGTAACCTT TTTTTTTTT TTTTGAGAA
TGGAGTCTTG CTCTGTCACC CAGGCTGTAG TTCAGTGGCG GGATCTCGGC TCACTGCAAC
CTCCGCCCTCC CGGGTTAAC CGATTCTCCT GCCTCAGTCT CCCGAGTAGC TAGGATTACG
GGCACATGCC ACCACGCCCG GCTAATATTT TGTATTTTA GTAGAGACGG GTTTCACTGT
10 GTTAGGCCAGG ATGGTCTCCA TCTCCTGACC TCGTGATCCG CCCGCCTCGG CCTCCCAAAG
TACTGGGATT ACAGGCCGGA GTCACCACGC CTGGCCTTT TTTTTTTTT TTTTTAAAT
TTGAGTTGA AGGTTGGGC AGAAAGAGAT CAGGATTGC ACTGCCCTGT CACATGCAAT
CTCCCCTGTC AGAGCATTAC TCTCAAACAT GGAAAAACTT TAAAATACAC AACTCTCCAG
15 ATGGACACAG CTGAATCATT TTCAGGAAGG CTCTGCCTAT AAAATTACCT TTGGCCTTAA
TTCAGTAATT AAAGGCACCC ACAGGTCTGA GCCCTCATCT ATGAAGGTTA AACGTATCAT
TCCTATCAGC ACTAATTGGT TTTATAGGAT ACAGACCTCT AGTTGCTGA ATAAACTTTG
GAAGGATTCA GATTATGAGG TTCTAAACTG TAGGCCTTT GAGGAGAAAG GTACCCATT
20 TCTTCTCCGA ATAGATCATT TTGTGTCTTC TCCTGGCTT AGCACATTGT CTTCTTAGG
ATCTGATAGT CTGTTTTAT ATCTTTTTGT TGTCATACCT TTTATTCTTG CATTCCCAC
TTTGACTAG CATTTCGCCT TTTCTCCGTT TCTGGAAGCC TGTAATTTC AACATTCCCC
ATTTCTCCTT TTGTTCACTG GAAAATTTC TTCAGTGTG GGATGCTCTG GAAAGGTACT
25 GTAGTTTGG GGGTCTCCCT GCCTGCTGGG CTGCATAGCA TATCTCTT TTTTGAGACA
AGGTCTCACT GTGTTGCCCA GGTGGAATG CAGTTGTGCA ATCATAGCTC ACTGCAGCCT
CCAACCTCTG GGCTCAAGCA ATCCTTCCAC CTCAGCTTC GCCGCTCAA TAGCTGGAC
TACAGGCAAT GCACCACCAT AACCAAGCGAA TTTTTAAAAA TTATTTTTGT AGAGACAGGG
TCTCACTATG TTGCCAAAC TAGTCTTGAA GTTCTGACCT CAGCCCCGCA AAGTGCTAGG
ATTACAGGCG TGAGCCACTG CTTCTGCTTG CATTATCTCA AATTTTTAGA GCCTAACTTC
30 ACAGTCTTGC CTCGGAGCAG GAACTGTTG AGGCAACAAG ATGGAGCCTC TGGTTCTTC
ACTAGGCAGA CTGTGCTCAA ACTGGGTAGC ATGAAAGGAA AGCAGCACAA TTAAAATTGA
AATTGGGGGA TTCTTCCCC TATTGAAATC CAAATAATT TCCTTTATTG TGCTTTTTT
TTTTTTCTTT AGGACCACAA GAAAAAAAAG TAGTGGTGTAA CTTTCAGAAG CTGGATACAG
35 CATATGATGA CCTTGGCAAT TCTGGCATT TCACCATCAT TTACAACCAA GGCTTGAGA
TTGTGTTGAA TGACTACAAG TGGTTGCCT TTTTTAAGGT TAGTTTTGTT GGAAGTTGGA
TTTACATTG CAATGATTTG ATATCTGAA CCTCTTCTGA TTAGTAGACC CTCAGAATT
TAATTTAGA TTAAGAAGAT GACCGGAATT GACACCACTC TTCCCAAGAG TAGTAGTAGT
TGGTATAATT TTGCCACTTT ATTAAAATT ATGATTTAG TAGTTTACA AAATACCAGT
GCTACATTG AATATGTATA AATTATGTTT AAAATTACA TTTGGTAAG TATGACTAA
40 TTCTTAATT ATTTCCTTA TTACTACCAC TTTATTCTA AATGTTGCCA TAGTCATTG
GCTTTGTTCT AAATCTGTAG GAAAGATAGA GAGATTACAC ATTTTGTTT CTTGCAGTTA
CTATGCTGTC CTTCCATCA CTACCTGTTG GCTGAGGTAG TGATAGGCCT AAATGATTCA
TTATCTTAAA TGTACTAAAT ATGTTGAGTA ATTTTTCTT CTAAACTAAC AGAAAGAGAG
AACCTAGGAG TTACTCCCTT AGGCTGGTTA AAGTGAAGG TAGCCAAGTC AACCCAGCTT
GTTTCTTCT CTCATTAGGA AAGAACTATT GTTCATTCTC ATAACACACT TTTTCCAATT
GCAACATAC TCAGGGTTAA AATAGTTAG CACAAATTGC AGCCCATTC ATTTGTTCTT
CACAAGCTGG AACTTTCTT GTAAGCTAAA TATTAAATGG TTCAGTAAA TTGGATACAT
AAGCCTGAAA CTAGGCGTTT CTCATTATAC ATAGAGTATA AATTAAGACA GACTTTTCA
20

TGGTGAAAGG TTTACAGCCT TTAAAACATC TGGGAAGAAG TGGGAAAGTA GGGAAATAACT
CTGTTAAATA TGATAAAAAGA CAAAGCACCA ACAAAAGGCCT AGTTCTAAAC TTGTTATAAT
TTCTCATGGG AGTTGTGGTT TGTACAAGG TTATGGCGGT CCAAGCAAGT TTACAATATT
TTTTAGAATA ATAACCTCCC AAAATATTT TAAAATAAGG ACCTTTCTTT AATATGGAAA
5 AAAAAAAGAT GAAATAGAGA GGAAGAGGTT GCCTTCCTC AACTGGATTG GCTGGTGTGA
TATGCAGCTG GTGGAAACCA GTTTGTCCAT GCTAGCTGTT GATGCCACC AGTCCACTGA
ATTAGTGGAA ACCCATGCTT AACCAAGTTGT TAACTATTT TAGTATCATT TCCAGATGCT
ACGCATTTT TTTGTAAAAA ACATACTTAC AATAGTATCT ATATAAATT TTAAAATTGT
AATGTATATG TGGTATATTT AACCTGAAAA TAATCTTGG TGTATACGTT GATGAGCCTA
10 GCCCTTGCA GACCTCTTAC AGTTACTCTG TCACAGCCCT TCAAAGGTTG TTTGACTTCA
GATTAAGAAT CAATTGCATG TGGAATGCAT GTGCAAAGGA AGAGATATTG AAGGCAATTG
TTACCATGTA CCATAAACCT TGTACATAAT TTTTGTCA TTTCTTTCC TCTGTCTCTA
TCCCTCTTC TTTGACACAC AATAGACACC TAGTCAATT ATTACAAAAA AAATGAATGA
ATGAAGTGGG ATTCAAGTTGG GAAATAGGTT AAATAAATT TAATAGGAGAT GAGGAATAGG
15 TAAAAAGAGA CAAGTACATA GTTTATTCTT TTGACTTAGA AAACCTTTGA TTCTTAAATT
CTGCAGAATT GGAGAAACTG GTGGGGAAAC TTCTAAAATC ATTATTTAAT TACCAAGAGAT
GTAATAGATA TAGACAAAAG CAGTTTCTT CTTTTATTAT TTTTCATCA GTTAGTTCTT
AGCTTAAATA GTAGTCCAAA GCTGGTAGGG ACAGAGGGAA TTAGCTGGTG GCTGAATGAG
GAATTGTATC ACTTTTGTG AATCACGGTG TAAGCACATT TGGTGTTTG CCATTGCCTA
20 AGAACATTAG TCACATTAGG TCAATAGAAA ATCACTTTT AAAGCCAAAT AAAGTTATAT
GTGTTCCCAA CCATGAGTTG GAAAGAATTA ATATATATGC TGTTGGAGGG TAGAACCCCTG
CCTAACATCA TGGTTCTGGA TGGCATTGAT CGAACCTTA TTCTTCATT AGGAATAACA
ATAGAAAAAA TACTCCTGCC CTACTGATT CAGGATATGT CTATTTAAA GTGCCATT
GACAAAACCA TTATCAGGGC CATGTTTCT TTTCTGCAG AAAATCAAC CACTCTGGTC
25 AGTAGTTAGG TCTTATGACA AGCACCAATAA TTTCCTTAGG CAGAGTAGAA TATAATAGGA
TACTTCTTT TGAAACTTAA TATAATCAGG TAGTTCCAGA TAAACATAGC TTGCAAAGTG
ATAAAATACC ATGTTATTT AGTAAATCCA ATTGCAAGAG TGATGGGAAA CAGAGTTAA
AAACTTAAGA AAGATATTAA GATGGAGTTG ACTTTGAATA ATAAAGTCAT CCACTGTTGA
TGGGTGACAT TTATTAATAC AGAAAGTTA CAGATTTCAC CATAAGCATC AGGGTATTTC
30 CTGCAGCTGG GGAAACCGTG CTTGAAAGGA TGCCTAACTC AAGGAAAACA CAAGCCATG
TAATAAGTAT TGCATGTGAG AATTGTTCTT ATAGAATTAG AAAGCATCTT TTACATTAAA
ATTTATTTT GTAAAAAGAG AAAACATCAA AACTTGAGTA GTATTTGCTA TTCAGAGT
CTTACATAAA CGAAAACATA CTCAACCTAC TGCCATTACA GAAATATTG ACAATTCTT
GCCATGCTTA CCTGCCATCG TTGTTGTTAT CCTACAAATC AATTGGATTG TCACGCCTCT
35 CCACTGACTG GAACCCTACA ACTTGCTTCC TTTTATCCTC TTTATATATG CTTCAGATAT
GCTTGAAGTA GATTGTTCTT TATTGTTCTT GCTGCTCGAG TTTGTTGAGT AGTTGGTACA
TGCAGAGTAT TTGTATGTTA TGACATATAT AGGTTTGTC CATGGTTCCG GGCTCATAAC
TCACTCCAA GACCCTGTT ACAGAAACCA GAATCTCTCT CTCTGATCTT CTCCTACCC
CCTTCATCT GCCCACTGCA GAACTCTAAT CTGATTATGG TTTCTAAGAC CCTCATAACCA
40 GAGAGTATTC TGCCCCATAC CATAGCAGAA GGAACACTGC ACAGAGACAC CAAGAAGAAT
CTGAACAGAC AGGCCTGTT AGGTTAGAT CATGCTCTTA TAACCTAATT ATATTTAAC
ATGGTTATCC ATGCTTTAAT CATGTGTATT CAATGAACGC TCCATAAAAG CCCAAGAAGA
ACAGATTGA GGGAAATTCTG AAGCGCTAAA CATGTAGGGT CTGACAGGAA GGTGAAGAAG
AACTCATCAT GCTGGAAGGG TGGTCCACCC TAACTCCGCA GGGACAGAAG CTCCTGTGTC

TGGGACCCCTT CCAGAACTTG CTCTATGTAT CTCTTCAAGT GCCTGTGTAT TTATATCCTT
5 TCAAATATCC TTTGTAATAA ATCATAAAC A TGTTTCCCTG AGTTTTATG CCACTCTGTC
AAATTAATTG AACCCAAAGA GGGGGTTATG GGAACCCAA TTTGAAGCCA GTCCATCAGA
AGTTCTGGAG GCCTGGACTT GAGACTGGTG TCTAAAGTGG GCAGGCAGTC TTGGGGACTA
AGCCCTCAAC CTACGGGATC TGAAACTGTC TCCTGGTAGG TAGCATTGGA GTTGTACTGG
AGGGCACTGA GCTGGTGTCT GCTGCAGAAT TGATTGTTG CTTGCTGGTG GGGAGAACTT
CCTACATATT TTGGGGTCAC CGAAGTCTTT TGTATTGATT GTTGTGCTG AGCTCCTATT
GCTAGAAAAT TGCTTTATTA AGTTCTGTTT ATTCAAGGGCT TATCTGAAAG AGAAACATTT
10 TTTATGATTT GAGATTCTA AGCCATTAA AAACTTGGTT TTAACAGCTT GAGAAATTGG
GGGTATAGTA GGGATGGAGA TACATATATT TGGATGTGAC TTCAAGCTAG ATAAAAGTTT
GGAAGGAATA AAAGTTGAC ATCTAAAGTT TTTGCATAGT TTGAGTTGAG CAGGAGGTAC
TAGGTATGTT TTTAAAATAT TTTTTTCAGC CAGGCACGGT GACTCATACC TGTAATCCCA
15 GCACTTTGGG AGGCCGAGAC AGCGATCAC CTGAGGTCA GCGTTCGAGA CCAGTCTGGC
CAACATGAAG AAACCCCGTC TCTACTAAAA ATACAAAAAT TAGCTGGGTG CGGTGACACA
TGCCTGTAAT CCCAGCTATT CAGGAGGCTG AGGCAGGAGA ATTGCTTGAA CCCAGGAGGC
AGAGGTTGCA GTGAGCCGAG ATCACACCAG TGCACTCCAG CCTGGGTGTC TCAAAATAAT
AATAATAATA AAAAATGAA AGATTTTTTC TTACTCAGCA TCCTCCAGGC ATTTTATTAT
20 CTGAGCACTT TATGGGAGTT GCATATTACA TTTAGGGCCC ACTCAGGTGG GTGGGTATCT
AAGCATTGTA AATAACCTTA TGTAACCTAA TAAGGAGTAA TCAGGCCTGT GGCAAGATGG
AAACAGTCTT AGAGGCATTC AAATTCAAAT TTCCTTTAAA ACACGGGCT GGCCAAAACA
AAAGACAATA CTATCTACAG GCCAGTTCT AAGATTATCA GATTTTAGTA GCATTTACCA
25 TTTCATTGTA CTTGGCACAC TTTAGCAAAT TTGCACTTCT TAAAAGTACC TGCAGGCAAT
CTCCTATATA AAAACACAAT GCAGGCTAGC TTGGCTCCTG CCTTTAATTC CAGCACTTTG
AGAAGTTGGA GACTAGCGTG GCCAACGTGG TGAAACCTCA TCTCCACTAA AAATACAAAA
ATTAGCCAGG CATAGCAGCG CACGCCTGTA GTCCCAGTTA CTTGGGAGGC CGAGGCAGTA
GAATCACTTG AACGCTGGAG ACAGAGCATG GAGTGAGCTG AGATTGCACC ACTGCACTCC
30 ATCCCTGGGTG ACAGCGTGAG ACTCTGTCTC CAAACAAAAC AAAACACACA CACACACAAT
GTAACAACAC GAAACAGAAT ACTGTGAAAA TGCTTAATTA TGTCTGACTT TACATGATGG
CTGGATATGT GATTATTTTT TCTCTTTTAT GCTCTTATGT ACTTTGTTCA TTTTTAATGA
TGATCATGTA TAAAGCTCCT CTGTGTAGCA TTCTCTCCAC CAAATTGCC AGAGACAGGA
AGTCCTGTAA AACAAACTAA GCTCCAAAAA ATGACCTCCT GTTGAATAGG CTTTTTTTTT
35 TTTTTTTTTT TTTTTTTTTG AAATGGAGTC TAGCTCTGTC TCCCAGGCC TCGCTCCTTC
CACCTCCTGG GTTTAAGAGA TTGTCCTTCC TCAGCCTCCA GAGTAGATTG GATTACAGGT
GCCGCCCATC ACGCCAGCT AATTTTGTA TTTTTAGTAG AGATGGGGTT TCACCATGTT
GACCAGACTG GTCTTGAAC CCTGACCCCA AGTGATCCGC CGCCTGGAC CTCCCCAAGT
GCTGGGATTA CAGGTGTGAG CCACCAAGCC CAGCCTGAAT AGGCTTCTA ACCTACTGTT
TCTCATTATA CTTTCTCTGA GGCAGTAAAAA GAAACTGACT CTAAAAGGGA GCAGTAGAGA
40 AGGAACCTCAG ATTTTATTTT GAAGATTAAG CTACTCAAGG GCTGAGGAAA TATGTAGAGG
GGAAGTAGAC TCATTATGGC TGAAATTAA ATTGGTGTAGT AGTGAAGCAA ATCTTAGGGC
TTCTAATTG AGCTCTTGTG TGAAAGGCTC CAATCTTAAT AGAACTATAA GCTAAAAAAA
ATGACCATCA GTGTTCTAA AGCAAGTTGC TACTAAAAC AAGAACACTT TGGGAGGCTG
GGGCTGGTGG ATCACCTGAG GTCAGAAGAT GGAGTCCAGC CTCAACATGG TGAAACCCCT
CTATACTAAA AATACAAAAA GTAGCCGGGT GTGGTGGTGC ACGCCTATAG TCTCAGCCAC
CTGGGAGGCT GGGCAGAAG AATCGCTTGA ACCCGGGAGG TGGAGGTTGC AGTGAAGCTGA

CATTGTGCCA CTGCATTTCA GCCTGGGTGA CAGAGTGAGA CTCTGTCTCA AAAACAAAAC
AAACAGAAAC AAGAACTCTA TGTTGAGAAA TCCATACTAG AGGGTTTAAT TTCTCACTTT
5 GTGTGCTAGT GATTAATAAA TTTCAAGCTT ATCACACAGC CAGAAATGTC GCCTGTGTT
CTACAATAAA AGATTTGGGA ATATTGTGAC ATTTTTTCAC TGAGCCTTCT GGGTCATTA
TTAAGATATA ACAATTTAG AAGACTTATT GAGATAGGTA TACTTTTAA AATTTGGATT
CAATATATCA AGCTCAGCAT TTTGTCTTT TTTGTGTTG TTTAAACAA CTTTGGGTTT
ACTTGTAAAGA GTATTTCACT GGAGGAAAAG TATCAGAACT TTAGCACTGA GTGCCTGCCT
10 AACTATTTA TGTCTTCTT CATGATGATG CTTAACATCA ACATAATGAC TCAGTTAAC
ATTGTAAACT CTTGTAGGA TGTCACTGAT TTTATCAGTC ATTGTTCACT GCAGCTGGGA
ACTGTGGGA TATATGATTT GCCACATCTG AGGAACAAAC TGGGTGAGCT CCATTGAAAAA
ATTTCATACT TTCTTATTAA AAGCAAGCTT CCTCACTTCC TGATGGCTTA TGCAAATTAA
TGCCTATTAA TTCATCAGAG GGTGACTGGA GAGATTATGG TTTCTTATG AATTCTTGC
TTGAGCCTGC TTGTCTGCTG TTTCTAAGTT GTCTGCACTA TTTTATGTGA GTAATTTCT
ACTTTATTAT GTTCTGTTT CACATGTCAA ATCAGCTCTC CCAAGAAATGC TACTTGTAAAC
15 CTAAGTAGAC GTGAACCGAA AAGGGTAAAG ACCCAGCTAA AAAAAAGGTT GACTCAAGTT
CAGTTCACAT TCGTAAAGTA GTTAGGCTGC CTTAGGTTGC TTATTCTTT CTTGAAAAGA
ATCCCTCAAG AGAGAAACAT GTGAGGCCAC AGCAGCTTAG ATCTGTCTTC CACAGAGAAG
GTGGCTTTA CAGAGAAATT GACACTCA TCACTTATCT GACATGACCC AGCTTGAA
AACTGGCTTC TATTAAAATA GTCTTAATAG ACTATTCACT GAGGAGGAGA GAATCTTATT
20 CACTCTTAC ATTCTCTTCA CATTTCAGA ATAGATGTT AAATCATTGC TCACACTGGA
TCCATAATG TCTAAAATGT TGATGAAGAA ATAGGTGATT AGAGAGTAAA ATTAATAGAG
ACTTACCCCTT CCTTGCAATT TAACATAATA TTCTTCCCC TTTCCCTTCC TCTGTTACTT
GGCTCTTAA TACCAGAACT GAGATATGAA AAAGGGAACT GGGAAACAAGT ATTGAAAGCA
CCATAGGTTT ATCTTATATT AGCATTTC CAACTTTATA ATGAACCAGC AGTGAACCCAC
25 ACATCTGCCA GGTAGGAATT GCTCATCAGT TGTGTCGCGT TATCTGTTA AGCTTCAGCA
ATGTTCCATA CAGCCACTAA TAACAGATCA AAGTGAGCAT TAGGGTTGAA ATTAGTAAGC
TGTTTCTTCT CAGTTCTTTC TGGTAGTTGA ATAAATATAG AATGTATTAA ATAGTTTCT
TTATTCAGA ACTTCTCAGA GTCTGTAATA TATTGTATGG TGGTAGCTTA GGAGGAAGAT
GCAATAGGAA ACTTTCCC A GATAGGTTCA CTATTTTTT TTTCCACGAA AAATAAGCTG
30 TTCTCAAAAT ACAGTTTACA AAATTTTATC CTTAACCTTT CACTCTTCT CCTAGTTAGG
GAGACCGCTC CACCAGTAGA AAAGATAAAC CCTGGTAATT TGTTGTGAA ATGGGATAAA
TAGCCTAGTA CCTAGTCATG TGGATTCAAGG CAGCACTGAG CCTAAATTAA AGTTGCAAG
GTATACATGT TAATGTATCT AAGTTACTAT ATTTAGCCTG TTTCTTAAGT ATGTTTCAGA
AACATATTG TTTTTTCAG TGGCAGTTAC CTTCAGATGC ATGTGCTTCT AAAGCATGTT
35 GGTTGCATGT GCAGTACATG TTGCTTAAGC ATTTAGCTTC AGAATGGCAT CTTCCCTGTG
AATGTCTTAA CATTACAAA AATATACCAAG GATCTCAAAT ATCAGTGCTG GTATTTTTT
TTTTTTTAC TTAAAGAAAC TGATATGATT AAATATTAAG AGACAATATG ATCCTTGTG
GCTTGTAACC CTAGTTTTA TTGCTTGTA GTTATTAAAT AGAGCATCTG TTGAGGGACT
CTTTAAAAC CACAGCCATG AACAGACGTT GGGGCTAAGA GACAGAGCAG CCTGCGACAG
40 TGTGGACCTA CCTGTAGCAG CTAGCAAAGG CCTCTAGCAG CTACAGTCCC TTCTGGAGTC
TTTATTGCA TGCAAAATGC AAAGGAGTCC TGGTGACCTA CCTCCAAGGC AGCTGCCCTC
CTGAACACTC CCTTGAAAAA CAGTAAACAT CATTGAA TGTGAACAAAC CAGAGACTAC
ACAGGAGAAA GGAAAAAAA ATTCTGAAGA TGCAAAATCT TGGGTGGCTT CACCGTTCA
TTTTTAATA AAAGGAACAA TATACAACAC GTTGTCTTT TTCTCTTTG AAATCCCTTC

TATTACAGTG ATTTTTTCT AAGATTGTCA GGATTGAAG TGTATGTTT GTTTTATTCA
CAGCGTAAAT TTTATTACA GTTAACTGT CTGCCTGAGT GTCTTCCTT TCTCTAATTA
CCTTGAGGAA CCCAAGAGCC TGTTGTAAGG AGAAAATAAG GCCCTTGGAT CTCTTGAGAT
TCACAGATAT AAGTTATTGA AGGGAAGATG GTCCTATGGA GGACATATT AAAGAAGGGA
5 AAAAGAGGC TTTCTCAGAT ATGTCAGACT GCTATAGTAT ACTTCTACAG ATTATAGACC
TCCAGTACCT CTGGCCAGAA AGATGGTATC GTAAACACCC TATTTTTTT CTTTTCTTT
TTCATTAGGT ACAAGCTTG TGCTAAGAAG TTGACATACT ATAAGCTACA AAAGTTCTGT
AAAGTAGATA TAACTAGTTT CATTTTATAG ATAGAGAAAA TTAATCTCTT ACAGTGCTAA
10 GCTCACAGAG TTTCTAACTG TAAAATGCTA GAACTTGTCT TTCAAGCCTA AAGACTTCCT
TGGGGCTAAA TAGTGAAGAA AGCCATTCA CAAATAAGTA AATGGTATTT AGAGGCATAT
TTGGATTTC C TGGTAAATTC CAGTCTGTGA GCATCATGAA TATTAGTTA ATGTTGCATG
GGCTCATGTT GAAGTTTAA GAGAAGAACT GCCTTGAAGC TTAGGTTCC TTAGCTATTA
GGCTACTGAC TTTCTTGCC CAAACCAGGGT TTTTCATTG AAGACCAAAA CTTACCTTCT
CCTTCAGTTT CTAGTTGGA AATTGGTAGA AGAGCTTGT AAACCTCAA TTAAGTACAA
15 ACTAAGTGTC ATAGTCAAAT TTACTAATCT TAATTACAGT ATTGTTCAAC TGATTGCTAT
CTTCTAGCTC TTTCTGCCG AATAATGGTC TTGTTTCTG CTCTGTTGGT TTAGAGCTGA
CTTCTTCAG CTTTGGTAAG CCTGAAATT TAAGGGTTATG TTTAATTCA TATTGCTGGG
TGGACTTTCC TCTCTTGCA T TCTGTTGA ATAGAAGAAAT TTTTCTCTAG AGAGTAGTTT
GTCATCCTTA CTCTGTTGAT TCAGATGACT CTTGTATGA TCTGAGAGGT ATACTGTTCT
20 GCTATTCTGA GAAGAAGTAT TTCAGAAAGA TGAATTAAAGA GTACAGTGGA CTGCTCCCAC
CTGGAAACTT TTATCTATCT CACCTCTGGA CCTGATAAAAT TCTTTATCAC TCAGGACCTT
GATGACGCTG CTCTCTGAAA CCCTCCCCAG CTCTCTCTAT TACCGTGAGA AACATCAGAA
CTTTGGTTCC CATTGCATAT CGCAGGTACC TCTGTTTCA TGCCATGCTG TAATGGAGTG
ATTGGGTAGC ATGTTTTCAT CTCTTCCAG ATTGAAAATC TGTATTCTC CCTGTATATC
25 TTCAACACCT AATGCACATA GAACTTTGTA GGTACCTGGA AAATGCACCA CAGTTTCTT
TTCTTTTGC AGACTTTCA CAAGTATTAC CAACTTACAA AGAATTAATT TTGTAGGATT
CTAGAAAGAC AAATCAGGAA TGGTGCATA TACATTTTT TTGATTCCCT GCTCTAAAGA
ATATTATCAG GTTACCTTCC TGCAGAGTTT TAAAAGAATT GCATATTCA AGCTGACTTT
CAGGATGTA A TATAACCAA AGCAACTGAT ATGTAAAAAA TATATTCAAT GGCATTCCCA
30 GATTTTCTTC TAGGGTGT TATTGTTTG GGTACCTTACAT TTAAGTCTT AATCCATCTT
GAGTTAATT TTGTATAGGT ATAAGAAAGG GGTCCAGTTT TAATTTCTG CGTATGGCTA
GCCAGTTCTC CCAGCACCAT TTATTAAATA GGGAACTCTT TCCCTATTGT TTGTTTTGT
ACGGTTGTC AAAGATTAGA TGGTTGAGA TGTGTGGTCT TATTTCTGAG ATCTTCATTC
TCTTCCACTG GTCTATGTGT CTGTTTTGT ACCATGCTTT TTTGGTTACT GTAGCCTTGT
35 AGTATAGTAT GAAAGATAGC ATGATGCCTC CAGGTTGTT CTTTTGCTT AGGATTGTCT
TGGCTATACG AGCTTTTTT TGGTTCTATA TGAATTAA AATAGTTCT TCTAATTGTG
TGAAGAATGT TAATGGTAGT TTAATGGAA TAGCATTGAA TCTGTAATT GCTTTGGGCA
GTATGCCAT TTTCATGATA TTGATTCTTC CTATCCATGA GCATGTAACG TTTTCCCTT
CGTTGTGTC CTCTCTCATT CCCTTGAGTA GTGGTTGTA GTTCTCCTTG AAGAGATCCT
40 TCACTTCTTC TGTATTCTA GATATTAT TCTCTGTGTA GCTATTGGGA ATGGGAGTTC
ATTCATGATT TTGCTCTCTG CTTGCCTTTT GTTGGTGTAT AGGGATCCTG GTGACTTCTG
CACATTGATT TTGTATCCTG AGACTTTACC GAAGTTGCTT ATCAGCTTAA GAAGCTTTG
GGCTGAGATG ATGGGGTTTT CTAGATATAG GATCATGTTA TCTTCAAACA AAGACAATT
GACTTCCTCT CTTCTATTT GAGTACGCTT TATTCTTTC TCTTGCCTGA TTGCCCTGGC

CAGAACTCCC AATACTATAT TGAATAAGAA TGGTGAGAGA GGGCATCCTT GTCTTGTGCC
AGTTTCACG GGGATGCTT CCAGCTTTG CCCATTCACT ATGATATTAT CTGTGGGTTT
CTCATAAAAA GCTCTTATTA TTTGAGATAC GTTCCTCAA TACCTAGTT ATTGAGAGTT
5 TTTAACATGA AGCGATGTTG AATTGTATCG AAGGCCTTT CTGTGTCTAT TGAGATAATC
ATGTGGTTTT TGTCTTTAGT TCTGTTATG TGATGAATGA CGTTTATTGA TTTGCATATG
TTGAACCGC CTCGCATCCT GGGGATGAAG CCAACTTGAC TGTGGTAGAT AAGCTTTGG
ATGTGCTGCT GGATTTGGTT TATCAGTATT TCATTGAGAT TTTTGCGTC GAAGTCATC
AGGGATATTG GACTGAAGTT TTCTTTTGT TGTCGTATCT CTGCCAGGTT TTGGTATCAG
10 GATGATGCTG GCCTCATAAA ATGAGTTAGG GAGGAGTCCC TCCTTTCAA TTGTTGGAA
TAGTTTCAGA AGAAAGGGTA TCAGCTCCTC TTTGTACCTC TGGTAGAATT CAACTGTAAA
TCCATCTGGT CCTGGACTTT TTTTCATTAG TAGGCTATTT ATTACTGCCT CACTTCATA
ACTTGTTATT GATCTATTCA GGGATCCAAC TTCTTCCTGA TTCAGTCTTG GGAGTGTGTA
TGCATCCAGG AATTTATCCA TTTCTTCTAG ATTTTCTAGT TTCTTGCAT AGAGGTGTTT
15 GTAGTATTTG CTGTTGGTTG TTTGTACTTC TGTGAGATCA GTGGTGGTAT CCTGTTTATC
ATTTTTTATT GTGTCTGTTT GATTCTTCTC TTATTTTGA CAAAGCTGAC AAAAAGAAGC
AATAGGAAA GGACTCTCTA TTCAATTAAAT CCTACTGTAT ATCTGGCTAG CCATATGCAG
AAAATTGAAA CTGTTCTGT TTCTTAATCC ATATACGAAA ATCAACTTAC GATGGATTAA
AGACTTAAAT GTAAAACCCA AAATTATAAA ACCCTGGAAT AGAATATAGG CAATATCATT
20 CTGGACATAG GAATGGGCAA AGATTTATG AGAAAGACAC CAAAAGCAAT TACAACAAAA
GCAAAAATTG GCAAATGAGA TCTAATTAAA CTAAAGAGCT CTGCACAGCA AAAGAAAACTA
CTGTCAGAGT GAACAGGCAA CCAACAGAAAT GGGAGAAAAT TTTTCAATC TATCCATATG
ACAAAGGTCT AACATCCAGA ATCTACAAGG AACTTAACAA ATTTACAAGA AAAAAGGAGC
CCCATTAAAA AGTTGGAAA GAACATGAAC AGACACTTCC CAGAAGATAT TCATGTGGCC
25 AATAAACATG AAGAAAAGCT CAACATCACT GACCATTAGA GACGTGCATA TCAAAATCAC
AATGAGATAC CATCTCATGT CACAATGGTG ATTATTAAAA AGTCAAACAA CATGCTAGTG
AGGTTGTTAGA GAAATAAGAA CGCTTTACA CTGTTGGTGG GAATGTCAAC TAATTCAACC
ACTGTGGAAG ACAGTGTGGT GATTCCCTCAA GGATTTAGAA CCAGAAATAT CATTACTGCA
TATAGACCCA AAGGAATAGA AATCATTCTA TTACAAAGAT ACATGCACAT GTATGTTAT
TACAGCACTA TTCACAATAG CAAAGACATG GAATCAACCC AAATGCTCAT CAGTGATAGA
30 CTGGAAAAAG AGAATGTGGA ACATAAACAC CATGGAATAC TATGCAGCAA TAAAAAGGAA
TGAGATCCTG TCCTTTTCAG GGACATGGAT GGAGTTGGAA GCTGTTATCC TCAGCAAAC
AATGCAGGAA CAGAAAACCA ACCACCCACAT GTTCTCACTT ATAAGTGGGA GCTGAACAAT
AGAACACATG GGCACAGGGGA GGGGAATAAC ACACACTGGG GCCAGTCAGG GGGTGGGGGG
TCAAGCTGAG GGAGAGCATT AGAAAAAATA GCTAATGCAT TCTGGGCTTA ACCCATTAT
35 GCCTAGTGTGTT CCATTTCTGG AATGCTAACG ATGTGGAAGT TCTTTATATC CTGCTCAAGG
TCATTGCCAA GGTCTGATTT TTCACATTCA ACAAAATGCA ACCTCTGGCA TAAATGGGTT
AATACCTAGG TGATGAGTTG ATAGGTGCAG GAAACCACCA TGGCACATGT TTATCTATGT
AAGAAACCTG CACATCCTAC ACATGTACCC TGGAACTTAA AAAATTAAA ATATATATGT
40 ATATATATTT AATATGGAAT TTTAAAATT ACTAATGAGT TCTTTATCT GAGTAATT
GCATCAACAT GCTTTTATTA TGGAAGAGAA GATTCACTGA GTACAAAATT GCAGATACAT
GTGTCAGAAG ATCCCTGAAT ATAATAAGGC TTAGTATTCT GTGTCATAAT TGCCTGTTG
TATTCCCTCTC TGGTCTTTAA ACTTCATTAG GGCAAGGATC AACTCCATCT TACTAACCAT
TTGATTCCT ATGTATTACA CGATATATGA CCAATAATAA GCCTTCATAA AATACTTGTA
AAATAAGAA TGTTATGTAA TATATCATGT GGTATTGTTT TATTGATGTG TTCTTTGAAG

TGTTACCTTT GTGTCCTTAG AGGCTATGGT GGGCCCTCAG GCCATATATG TTGGATGATT
 TAGTGAAAAA GAGATTAGAC TTTTATATT TATTCTATT ATTATTTAT GTTTTAATCT
 TGTGTAAGCC TATAGAGGCC AAGATAGGGC TCAGAGTTAA AAACAGTGAT GTTCTAACTT
 5 TCAAAACTAC TAAAAACAAA ATTAGCTATC AAGGGAAATC CTGAGTGTTC TGACTGCAGG
 GGTTCCCCAA AGCCTGGAGG ACCGAAGCTT TGAAATGCTG CATGGTGTGC CTCACCTCTG
 TTGAGTATAG TGAGGAACAT GAGCTGAGGA GTGGATTAGA TGAACCTTA AGATTTTTT
 CCAACTTAC GGTTGAGTT TATGAGTAAG TAAGTGAGGA ATGAGCCCAG GGACTTTGAC
 TACTTGCTG TGCACTAAAT TGCTGTTTC TTAGCTAATT CTGGAGTGTG TATTTGGTT
 TTAATGTAAT GTCCTTCAG CACTGGAAAT AAACAGATTT ATGACCCCTCC ATTTTCACAT
 10 GTAACCTCCA GTTCCCTTCAG TAAAATGCTG GAAGTTGCTT GGCTGTAACC ACTGTAAC
 TTTCTGGCA GTTTGCAATT TTATTACAAT TCGACTTTT AAATGCATGG TTCAGTCCAG
 TTAATAGGCA GAAAATATA GCTATTCTT GGAACGTACA TAGGGAAGTT GGTTCTTTC
 TGAGGAAAGA TGTGACACTC TGATGGATT GGCAAAATAC CAATTCTAGA AAATTGCT
 15 GGTTCTTTC TGACAGCAA TGGAGGTAAT TGTACGTATA TAGATTCTA CAGTAAACAA
 AGGGCAAAA TGGAAACATT TTTGTTTT TAATGTTGCT GGAATATAAT ATTTCCATAT
 TTGGTTAACATC ATCAAATCAT TCAACTCCAT GCAAAGTACT CTTTCAAAA TTACAAATCA
 TCTTGATGGA TTAGACTTGG ATGTCTAAAT TTTTATTAT GTATTCTAAA TTATCTATAG
 GTACATTTG AAAATTATAT ATACATGTAT CACTGTACTA CATTGTTT AAAATGTGAA
 20 ATATTAAGCG ATTAAATACT TTATTAATGT CATAAAATTG TGCCCTACCC AAAATATATT
 TTAAGAAGAT ATCATTGAA AATAATGTAA AATATTCTTC ATTATTATGA TTTGTTGCC
 AAAAGTAATG TTAGACTCCC TGTTACTGAT AATTTCTATG ACTGTAAATT TTCAACATTT
 GGAAACAAA TTTCTGATG TGTTTAAAAA TCTTAAAGAA TTTCAAGTAA TTATCAATTG
 TTTTGTAA AAGTATGAA GTGTCACAGC TTTAGAAGGA TTTAGAATAA TTATCAATTG
 25 GTATTAGGGA TGACAATTAA AATAGACAAA TTTTAACTGA AACATTGCT TTATTGATGT
 CTAGGACTAA AAAAATATT ATAGTGTGG AACATTCA AACATCTGT TTTAAAAAC
 TTCTGTCAAC ATGTTGACA GAAGTCAGAA TAGGGTAATT ATAAAAAAC TGCCATTTC
 TTATGTTGC ATCATTAGTA TTTGCAAATC CTTGTTCTT TGCAAGTATA TTTAAACTA
 CTTGTCATT TTATGAGTGT TAGCCTGATT AATTCAGATA AGATGCTCAA GTCAGTTAA
 GCAGGCAGAT GTGAAAAGG GAATATGCTG TGCTTAACATCTTATAGG GACTTACTAG
 30 CACACTGTAC ATAACAGATG AACATCTGAT ATTAGATCTG ATGGGCACAC CAATGTTGTT
 TCATATATTA AATAATAGTG GAAGTTTCA ATTAAATTAG TTTATTATT TATTATTTA
 TTTTGAGCC GGTGTTCCC TCTTTTGCC AAGGCTGGGG TGCAATGGCA CATGTTGCC
 GCTCACTGCA ATCTCCACCT CCCAGGTCA AAGGATTCTT CTGCCTCATC CTCCCAAGTA
 GCTGGGATTA CAGGCGTGTG CCACATGCC AGGCTAATT TTGTATTATT AGTAGAGACA
 35 AGGTTTCACC ATGTTGGTGA GGCTGCTCTC AAAATCCGA CTTCAAGGTGA TCCACCCGCC
 TCAGCCTCCT AAAGTGTGG GATTACAGGC ATGAGCCACT GCGCCTGGCC AATTGTATT
 TTATGTTAGA AATATATATG GGAGTTACAA TACTTGTGTT CCAATGTTCC GCTGGAAACC
 ACTGGTTGAA AGGATAGGTC CCAGTGTCA CGGCCAGTT GGCAGTTTG GATAGAATAT
 AAGTTCTATG AAGGCATGGA TGAAGTATCC TTCATGGTAA TGATGGTAAT GTTGTATAG
 40 TCACCTGTT AATGTTGTAT AACTTTAA CTCTTCAAC TTCAATTCA TCTATATCTT
 TCATAGCAGT CCAGTAAAGT AGTATCAATT ATAGATATCA TTCTACTTT CAGATAAAGT
 AACTGAGGCT CACTAGTTT GACTTGCTCA AGGTTACGTA AATGATAAGT GACAGACTGG
 ATTTAAAGGG CGGAGTCCAC TTTACTAAAC TGCCCTGTGTC TTCTAGCCAA GTTTGTTATG
 TAAAGTAAGT ATTCAGACAT TGGCTAAGTA TATTATTAAT CAAGAAAACA GTAGAACAGG

TTTAAGGAAT TTTTTTTTCC CTCTCCTCAG TGTCAAAAAG CAGTCAGTA AAAGATCCAA
 GATAAATAAC TTCAGTTCCC AAAAATAACT TCAATTCCAA AATTTCCCTA TGACTAGGTA
 TATCTAATTG GATAGCTGAT AAAATGAGTG GGAAATTGAA GAAGGGATAT TTTGAGAATG
 TGTGGCAACT TAATGGTAAT ATTGAAGAAG CAAATGGATA GTTATTGAAT CATAATAGTT
 5 TTTTAAACTT AAATACTAGG GCTTGATCTT CATATTTAG GAAATTATCT ATTTGTTGCA
 GTAGAAATTA GCATAAAATGA ATGGAAACTT AAATTATCTC ATCTGTGCAA GGCAGTCTTC
 AGAGACTGAT GTTTATATAA GTGAAAACTC AGAAAAATTG TTTGGCTGCA TTTCAAGCTG
 GATAGCGGGG TCAAAACTTC AACTCTTAA TTGGTGATAC AACGTCAGAA ATCACAAAAG
 10 AAGTATTCA CAGGAAAAAA TGACATACCC CGAAAACATT CAAATCAAGA AAATCACAGG
 GGTACAGTGG ATGGTCCCCC TTAAGGTACT GCGTACTTGG CTTCCCCGCT GGTAAATGCC
 TCTCTTTAC TAAGTTGTG TCGTGCAGAG TGGCATTGCA GCTCTGTACT TTTACAGAAT
 AATTTCAGGT TCTAAAATTG ATACATTATT TCAGATTGTA AGAATAGTAT AAAATAAAAT
 TTTAAAAGGG AAAATAATTC CATCAATTAA ATGGAATGTG TAGGGTTAA GTTATAACAA
 15 CAAAAACAAA GTTGTAGTT TTTAGGATTA CTCATATAAA TAGTGTGTCT ATTAAGAATT
 ACTGGCTTTA TGAATATTAA ATAAGAAGGC TGGCCGTGGT GGATCATGCC TGTAATCCCA
 GCACTTTGGG AGGCCGAGGC AGACGGATCA CCTGAGGTCA GGAGTTCAAG ACCAGCCTGG
 CCAACATGGC GAAACCCTGT CTCTACTAAA AATACAAAAA TTAGCCGGGC GTGGTGACGC
 ATGCCTGTAA TCCCAGCTAC TCAGGCGGCT GAGGCAGGAG AATTGCTGA ACCCGGAAGA
 20 CGGAGGTTGC AGTGAGCTGA GATCGCGCCA CTGCACCTCA GCCTGGGTGA CAGAACAAAGA
 CTCCATCTTAA AAAAAAAAGA CAACCACAGA TATAAGGATT TGCAATTAGA
 AAGCTTGAAG GGAAATTGTTT GGAACCTCTA GGTCAACCTA AAATTACAAA TAGGAAAGCA
 GCTTGGAGA GGAACATTG CACTCTGAA GTAATACTGT GATTTGGACT GTCAGACATT
 GTGATATATT GGCATCACCA CACCTCTGTT AATTATAAAC TTACCTATTTC TCAGTGCCTT
 25 GTTAACTGTT GAAATAGGAC TTGACATTTG TTGTTCTTCC TCTTCTGTTA TTCATTTATT
 CACTCATTAA GTAGTTAATT ACTTCCTAGT GCCAGATAGT TTTCTAGGT CTGAATAAAC
 AAAATCTA TTGTCATAGT TATGTTTTG TGGAAAGAAC AACCTGATAG ACTATCGTGA
 AAAATAATG AGTTAGTATA CATAAAACAC AGCCCATGAG ATATAGTCTG TAATTATCAT
 CCCTGCTTCA TTTATTTATT TATTATTTA TTTATTTAGA GAATGGATCT GATTCTGTCA
 30 TCCAGGCCGG AGTCAGTGG CTGGATCATT GCTCACTGTA ACCTCAAAAC ACCTGGCCTC
 AAGTGATTCC CCCAACCTCG GCGTTCCAAA GTGCTGGGAT TAAAGGCATG AGCCACTGTA
 CCTGCCATT GCTCCTGTT TAAAGATGAG GAAACTGAAA GTACAGAGAG GAACGTGACT
 TGCTCAGGAT CACACAGCCA ATCAGTGGCA GAGCAGTCTA GGCAGTCTAG GCCTGAAGGC
 ATTATTCTTT CTTCTGTT TGCGTCAAAA ACCCTAGGCA GGCATGGTAA AAAGACTGAA
 35 AGAGCCAAGG CAACATAGGT ACCTGAAGAA TGGGAGAAC CCTATGAGGA CTGCTAGAAA
 TTTTAGGCAG CCCTTAAAG GCCTTGGCA ACAATTGAA CGTGACTCAT GAGTCAGTTA
 ATTCTCTGCA CCTTTTTTTT TTAGAGAAT ACAATGGATA AAGCAATGAC CTTTGTAAACA
 TGTAACCTG TATTGAAATC CTTCTCTCTC CATTAAATTG TGCAGAGTCA CCTTTCTTC
 ATTAGAAAGA AATTATTC CAAAGAGATT ATCGTGAGGC TCAAATAATA TAGGTGGAAG
 CCCTTGGGT TTAAAAAAA GCCTTTCTG TCATTCTGG GTATCTCCG TCAGTGATTT
 40 TAGTAATTGT TTACCTCTCT GGCTCTTCGT GAGAAGACAA ATTGTAGGAA GAAGTGGAGG
 ATGGCATTGG GTAGGATGCT TAGTTCTAGT TCCACCACTT CCTTAGGTCT TTTTTTGTC
 CCCCAAGACT ATATCATTTC TGTGTTCTCT TGAACCTGTA AGATTATGA GTAATCTGTA
 CAATATTAT AAAAGTATGAA ATCTCTGCCA TTGTGGCACA AAAGTAGCCA TAGACAATGT
 ATAAATGAAT GGGCATGCTG TTCTCCAGTA AAACTATATT TATAGAAATA GGCAGTCTTC

TCCAATTTA GTTGCTCAC CCCTTCCTA TAGAACTCAG GTTTATCATA TACAATGTTA
CACATTATTC CTTTGAATAA ATTGTTTACC TAATACTTT TAAATTGGG CATTCTAG
GACATTCTAT TGTTTTATT TTTAACTTGT GTTATACTGC AGTGAATGTC TTCTGTGTG
TAGCATTG TTTGTGTATG TGTGTCAATA TCTATAAAAGT AGCTCTTAAG AACTCTGGAA
5 ACATTGTCAG ACTGCTTCCT CAAACAGCAG AGGTGGGAGG GAGCATCGGC CTAACCCCTAC
CTGGCCTGCC TTACCAAGCAT GCAACCAAGG AGAAGAGAAT AGGATCAAGA GCCAGTGCCA
GAGAGAAGTA CATAATACAA CTTTATGGAG TTTAGGTTGC ATCCTATAAA GTTACCATGT
TACACTGTT CGATCTACCA AAGTAGAAGT TTTACATAGT TTAGTATGAA GTTAATTTA
TTAACCTTTT CACACAACAC ACCCTCTCAT ATTGCTGCC CACAATGTCC TGGGTCTTTG
10 TTTCCATGGT CATCATATTG AGTAGGGTT TAGCAGGGAC TTGAGTAGTT GGGAACTGAA
CTCATGCTAG TGAAGCCATT CTTCACTAGA ACCAGACATA TGGGATCCAG AATATCATGA
ACAACCAGCT GGGTACAGTG GCTCACACCT GTAATCCCAG CACTTTGGGA GGCTGAAGGA
AGCAGATCAC CTCAGGTCAAG GAGTTTGAGA CCAGCCTGGC CAACATGGTG AAACCCATAC
TCTACTAAAAA ATACAAAAT TAGCCAGGCC TGGTAGCACA TGCCGTAGT CGCAGCTACT
15 TGGGAGGCTG AGGCAGGAGA ATCGCTTGAA CCTGGGAGGC GGAGGTTGCA GTGAGCCGAC
ATCTCGCCAG TGTACTCCAG CCTGGGCAAC AGAGTGAGAC TCTGTCGAA AAAAGAGAA
TATCATGAAT ATCACCAAAG ATAAAAGCAA GATGCGGATA TGAGTAGCCC AGAGTAACCT
TTCCTGTGCT ATATAACACA GTGAGTGGTG ATTCCCACAA TTCAAGTCCA GTTTTCTGG
ATTCTAAAGC CATACTTTT TAACTCTCCC CCTAAATATG ATTGAAATTC AGAGAGTCAG
20 ATGCAATAAG GTAATTTAA CTTTAAGTGT AGTTAGCCC ACACATTCT CTAAACTGGG
GGTTGGAAA CTACAAACCT CTTTTGTAA GTAAAGTTT ATTGGAACAC AGCCTAGCTC
ATTCAATTAT GTATTCCCTT TGACACCTTT AGGTTGAGTA GTTGTGACTG AGGTCTTCCA
GCCTTCAAGC CTAAAATATT TACTATCTAG TCCTTCAAAA AGAAGTTGC CAACCTCTGC
AAACAAGTGA AATAATGTGT ATTAGAGTAG AGCAGGAGTC CACAAACTAC AACCTCTGG
25 GCCAAACCA CCCTACTGTC TGTTTTGTA AACAGAAATT TTATAGAAAC ATAGTTGACC
CCTGTGCCAA TTTGTTGAG TCTGCTTTT GATGGCTTT GTGCTACAAG GGCAGAGTTG
AACAGTTGCA AGAGAACTAA GTGGCCTGCA AAGCCTAGTG GTCTTGCTG TCTGGCATT
ATAGGAAGTG TTTACCAAAC CCTGCATTTG ATCATGGAAT CCCAGAATGT TTATACCTTT
GAGGATTTT TTTCTACTT TCCATATTTT ACAGGCAACA GAACTAAAGC CTCGTAATAT
30 CTTGCCAAG GTTCTACAA TAGTCTTTT TATTAATTA AGTCAGAGCC TGAAATTGTC
TGTGCCAAG TCTCTCCTCA GAAATTAGC ACAGCATTTT CCCTTCTGT CTCCCACATG
CTACACAGTC ATCCTGGATC ATGTGCCCG CCTCCATGAG GAGATGTTCT AACATGAGGG
TGACTAGGCT CTGGAGCCAG AGCATCTTAG TTCAAATCCT GGATCTGTT AAAATAATAG
TTATTTTCA GTTCTAACAC TGAAAATAAGA CAAAAGGCCA TATTACTGAA ATAAAGTGT
35 AAGAATATAT TTAATACTCA GGTTACTTTA GCCCCCTCAT GATCAGAATA TGCTAATAAA
TTGAGAGATT GACAGAAAAA AAGGATTCTG TTTCACTGTG GCTCACACCT GTAATCCAT
CACTTTGGGA GGCGGAGGTG GGTGGATCAC TTGTCTGGAG TTCAAGACCA GCCTGGCCAA
CATGGCAAAA CCCCCTCTACTAAACAAAT ACAAAATTA GCCAGGCCTG GTGGTAAGCA
CCTTTAATCC CAGCTACTCG GGAGGTTGAG GCAAGAGAAT CGCTTGAACC TGGGAGGTGC
40 AAGTTGCAGT GAGCCAAGAT CACACCACTG CACTCCAGCC TGGGCAACAC AGCAAGACTC
CATCCCTGGA AAAAAAAAGT TCATCATTG TTCTGTAGT TTCCTAGTGT GGCCCTTGTC
TCAGTGGAAA ATTGTAGTGT GATAGACTGA GGAGAATTCT GTTCTGCCA CTTCCCTGT
TGCCTGGTA GGAGCTCCTT ATCTCTGAGC TGAGCTCCTT TTGTCCTTGG TAAAATGTGG
TTATTATCAT CTGCATCCTC ATAGTCTTCT TGTAAAGGACT GTGAAGTCAC TTTGAAAAA
18

TGCCTTGAAT ATGCCAGATG GTCTGTTTC TTTCTTTTT GTTTGTTTT GTTTGTTTT
CCCTTGGCCT TCATTCCTGG AAGTGTGG TTATATACAT TTTGTTGGT GGCAACAGT
GCTAACACACA GAAAAGTTG TATGTCTTCT TTTCTAGAA TAATTCTTAC AGTGTAACT
5 CTCAGGGTGA AGTTTTGTT TTGAAAGGGG CACATTTACT GTGAATGAGA GCCATGGAAA
TGGACCTGAA ATTGAAAGAC AGATTTACT TGGTTTCTT TTCTCTTATT TGTTGCAGT
ATAAAGAAGA GGGCAGCAAG GTGACCACCT ACTGCAACGA GACAATGACT GGGTGGGTGC
ATGATGTGTT GGGCCGGAAC TGGGCTTGTGTT TCACCGAAA GAAGGTGGGA ACTGCCTCTG
AGAATGTGTA TGTCAACACA GCACACCTA AGAATTCTCA GGAAAAGTGA GTTGCTACAA
10 ATGAGACATA CGTTAGTT TGTTTATAG TATATATGAA TATGTGTGTA CATTGTTGGA
ATTTTAGTT GATTATACAA AATATCTTG GCTTAGAAAT ATTAGGCATG CTATGTAAAA
CCTTACTGGA AAAATAAATT GACCAACATT ATTGAGAGTA TTTTTCAAA GTGTTCCAAA
AGTAATGGAC CAATGATTAC TTTAAATGAA ATCATGTAAT GGACCACAGA ATTGCAAATT
ACTAATAAAG AAAAGCCATT TTGCTTATTG CCATGTAATA ACATGTTGCA TGATACAAGT
15 AGATAACGTAT GTTATGCTG CAACAAGTAT AGGTGATACT AATTGGCAA CTTTAAACA
AGACCATAAA TAACTGAAAT CAAAGTTCTT AGTATTTATG CAGCCTGTTG GTTTGCGAGG
GCTGCCATAA CAAAGTACCA CAGACTGGGT GGCTTAAGGA AGGTTGTTTT TTCACAGTTT
TGGAGGCTAG AAGTCCAAGA TTGAGGTGTT GTCAGGTTG GTTCTTCCA AGGCCACACT
CCTTGGCTTG CAGATGTCTG CTTGCTTACT GGGCTCCTAA ACAGTCTTT CTGATTGTCC
TAATGTCCTC TTCTTATAAG GGCATCAGTC ATATTGGATT GGGGCCACC CATATGACCT
20 AATTTTACCT TAATGACCTC TTTAAAGCTC TGTCTCCAA TACAGTCATA TGGCTGTGAT
ACAGGGGGTT AGAATTTCAA CGTATGAATT TTGAGGGGGA CACAATTCA ACTATAAAACT
GCAGTTAATG TTTACTGTTA AATTAAATCT ATAACGTACT AAACACACAA AACAGGAGAT
TTCAAATTAG ACTTTATTAG TTTGGAAGG AAGAGATGAA ATGTTGTTAC TTTTCTGTT
TTATGTAGTT GTAAAAGCCA CTAAAATGT ACATGTCACAA ATCATCCCAA GCCAGGCAAC
25 ATTGTAATG AATGACTGCA GCACAGAGGT AAGAGAGATG ATTTAAAGGA GGATGAATTG
TCTGCAAAGG TGCTGGCAA GATCTAGACT AATTCACTCCC CTTCATTTA AGTTAGGTT
TTAAATAGCT TTGTTGGCT TGATCCTAGA GCTACACATT TACTTTAAC TTGTTACTT
TGTACCTAT CATTAGGAT ATGCTATGTA CTATTGACT CTATTGATA TTTCAAACAT
TCTCTCATT AGTGAAGAAG CTCCAACCG GAGTGCTCAG AACCTACATT GCCTCACCTG
30 TGAAGTGAGC ATGTGGGACT GAATTGCCTT TGAGGTCTCT TTCAGCTCTT AATGATCTGT
TATCCCATAG TAAGATAACAT TATTGTTAAT CTCGTTGGAT CTTAAACACC AAAATAATA
GTATTAAAGA CGTAGGATGC TATCTTGTCA TTATTTAAT GCACATGTCA ACACACGAGG
TTTGCTAGA TGTTATGATA AAGCAAGCGA AACAGCAGGC TACTGCTCCT CAAATATCTA
CAGTCATGA AATATTGACT GCCGTTGTTG GAAAATATCT CAAAAGTATT TTATTTACAA
35 TTTAAATAT ATTCTTTGC CTAAGAAC AATGAAACAA CTATAACTTG TTTTGTAT
TGTGTTAGTG TAGTCATACA TATGATACAG TTTTATAATC TGCTTGATAA GGGTACGTCT
CTGGGTTTT CACACTTGCA AACAAAGTGGG AAAGAACTTC GACAAAAACA TTAAACATT
ATCAACCTGG TTGCACGTT TTCTGATTAT TTTCAGTGCT CTTAGGTTTT TAGACCATAT
40 CAAAATTCTC TCTCTACACA CATGATTTAT AAGGGAAAGA AAAGTTGCA GTATGCCAA
AATGTCCTC AGGTTAATTA TATCTTGTGTT TGCATAGTTG AATGTTATG GAAGTCTTAG
ATAGTGAGTC ACTCTGAAAC CACAGACTCT TTTTAAATAG TAATAATATG GTGAGTTCTA
TAATTCTAG CATCTAGCAC ACAGGTATGA TAAATGTTT AATTAAGTGA ATAATCAGCT
TCCTGAAATT TTCTACTTTG TAAATATGCT CTTTGTGTGTT TGAATCTATT TACATATATA
TCCAAGCATT TGGCACAAGG TAAACCAATG TGGCTCAGAT TTTAATGTTA TTGTGAATCA

CTTGGGGATC TTGTTTAAAAA AAAAAAAA AAAGGAGATT CTGATTCAAG TGATCTGGC
 TGAGTACTTA CAGTATGCAT TTCTAATAAA CTTCTGGTG TTGCTACTGA TATTGGTTCA
 AGAACTAAAT TCTGAGTAAC AAGTATAAAC CGAATGTCCA TAAAGTATAA TTTCTTTAG
 TATCATATGG TTATAATGCT TGCACTTTGT ATATATGTTG CATACAGGTT TTGCTTTCT
 5 CTTTTTCTT ATTAAAAAAGG TTATAAAGGT AACTGTAGA AACTGTAGGG AGTAGGGAA
 AGCTGGCAT CTCATGCCAC CATTCAAGAGG TAACCATGTT TGGCATTTTA AAAATGCATT
 TATAGTGTGT GTGTGTGTGT GTGTGTGTGC GCGCATAACAT ATATATGTAT
 GTATTATACA TATACTTAA TATACTATA ATATACATAT ATTACTGTGC AAAAGTTGAT
 10 AAATTGATTA TGATTATGTA TAAACTTGA TATCTTCTCA TATTTAATAT AATATAGTTG
 TGAGCACTCT TCCATATCAT TAAATAACTT TCTAGACATT TTAAATGAAT ATATAGAAA
 TGAAACATCA TTAATCAGTC CCTTATCTTA GAATATTAT TTCGGTTATT TACAATCTAT
 TAGTAGTTA AAAGTACTTC ATGGAAAATC CTTATACATG CATCTTAGTT AACTTGAACT
 15 TTTATTTTA GAGAAGACAT GCCAGGCTGT TCTAGGCTAT TGATACATAT TGCCAATTAC
 CTTCTGAAA GATTGTACCA CTTTCCACTT AGGCACAGAG TGTGAATGCC TGTTTAATT
 ACTCTATTAA CATCAGTAAA GACTATATCT GAAATTATCT CATTTTGGG GACAGGTATT
 CTAATAGGCT CTACAAGTAT GATCACAACT TTGTGAAAGC TATCAATGCC ATTCAGAAAGT
 20 CTTGGACTGC AACTACATAC ATGGAATATG AGACTCTTAC CCTGGGAGAT ATGATTAGGA
 GAAAGTGGTGG CCACAGTCGA AAAATCCCAA GGTAATCAAG CACACATTT ATCATTAAATA
 AAAATATGAA TGCTGAATAC CATCCTCCCT CCTAACAGT CCTACCAGTG TTGCTGCCA
 TTTTATTGGT CATCCCAGTT GTAACTTTT ATGTGATCTG TTAACGATCT TTATCTCCTA
 TTGATACATA TTCTGTCACT CTCCAAGTCC TATTTATGAG TTTTTCTTTC ACATGTCTGT
 25 TGTTTCCATT GCCCCCTTTC CATTTCTAC ACTCTGGCAT GCAGCTTGT CCTGTTCCC
 CCATGCTATG GATTAGCCTC CTGCCGGTAA CCACCTGAGG GGTTCTTCC GCCTGCTGCA
 TAAAGAAAAA CCATGGCAAT ATAGTAGAGA AAGAGCTTAA GAGACACTAG GCTGCCATG
 CCACGTGGGG GATGGAGTTC ATATTCAAAT TATCTGTCC AAAGCTCATA GGTAGGGTT
 TTTCAAAGGC AGCTTGGGGG AAGGGGTGGG GGTGGCCAGG TAACAGATGC TTGCTGCTGA
 TTGGTTGGGG TGGGTGAAAT CACAGGGAGT TGAAGCTGTC CTCCTGTGGG CTGAATTGCT
 30 TCTAGGTGGG GCCATAGGAG TGGGGTTGCT GGGTCCAGGT AGAACACAGG GTGTCAGACA
 TGCAAAAATA AAATAAAATA AAATAAGATA ANGTAAGAT AAAATAAAAT AAAATAAAAT
 AATAAATAAA ATAATATAAA AATTCCCTG AAAAGATATT TCAAAAGCC AGTCTTAGAT
 TCTACAATAA TGATGTTATT TGCTGGAGTA ATTGATGGAG TTGCATGTCT TATAACCTCT
 GGAATAACGG CTGACAATCT CTCAAGTCTG CGCCTTAGCT GGACTCAGGT TCCTCTCTC
 CCCACAGCCT GACTGCCCTC ATTAGCTTCA CAAAAGTGGT TGGGTTTCAG GGCAAGGCC
 ATTGTCATTT AAACTGTAGC CGAAATGACT TCCAAAGTTA GCTTGGCCA ATAGCCCAGG
 35 AATATTTAAG TGGAAAGGCAA GATGGGGGAT GGGTTAGCTT AGCTCTCTT CACTCTCATA
 GTTTTCTCAC TGGTATAATT TTTGCAAAGG CGGTTTCATG CCTGCCATCT CTTTCGTCGC
 TACCTCTCCC AGTTCCCATT CTTAGCTGTT TTATGAAATG CTTCTAGTTT CATCCTCTTA
 TACCAAGTTC TGGGAGACTG ATTGAGTAA TAATAAAACT CCAGTTTCCC ATACAGCCGG
 CTCTGCGTGA ATTAAACTCA TTTTCTATTG CAATTTCCCT GTCTTGATAA TCAGTTCTGT
 40 GTAGGCCGTG AGGAAGGAGA ACCCGTTGGG TGATTACGAG ACTGTGTTAC TGCCCACTAC
 CTAATGGATA CATTAGCCT GGTATCCAAA CCCATCTAAT TATGACCATA ACTATATTAA
 TCACCTTGCT CTGCTTCAA ACGATATGAC ACACAATGAA TGAAAACCTT CATTTCAT
 CTTCAATTGT GCTGTTCCCT TTGCCTCAA TAGCCCTCAA CTTGCCTACA GTAACTGTAA
 AATTGCCAC CTAAAAAAA ATCTCAAAT CCTCTCTATG CTTTGATGTC CAGCAAAA

AAAAAAAAAA AAAAAAAAGT TATCACTATA GCTCTATTAC TCTCAACCAG AGGTGGCATT
 TTTCATCTTT ACCCATAAGC CCCATGGTAT ATTTCTTAGT TTGGGAAATT ATATCATAAT
 ACTTTGAATC TGTCTAGTCA GAAATTAGA TAATTTAAA AAAAATTATT TTTCTGAAAG
 TATAGCTGGA TCTAGCTTAG GATCTACATC TATCTAATAT AGTTCCCTAA CATNTTGT
 5 AATGCCACN GGNATAAGTC CTGTAATGCC ATACTTGCT TAGGATCAT GTGACTAAGG
 GGTAAGGAAT TGGAAAGCAA TGGGGAGCTA GCAGAATTG AAAAATTATA TAGTAGGTT
 ATTTTCCTT AATACATTGA AATAGCCTCA AATTCTCAGA GAATACAATG TTTAACCTC
 TAATATCTGT AGAGTCATG GCTCAAATT GTATTATTG AATAAACTAC TAATAGATTA
 10 ATAATTACTC AACTAAAACA CTTTGAAATG TGAGAGTTCC CTTATCTCCC TCGCAGGCAT
 ATGACAGAGG TGTGGCTTCT CACACCTTG GTTGCCTTG CCCTACCACC CAAACCCCTA
 GGGGAGCAT GCAGAGGGGC AGGTGCAGAG GCCATGGGG GTGCTTTGG GCTCTGGCCC
 CACAGCAGTG TCTAGGAGTG GATGTTGGAG ACTCCTGAAG CCCAAGTGGG CATGTGTTAC
 AGTGTGCTCT TTCAGCTTAG CCGTCTGCAG ATGGCTGTG TTAATCAGGT CATTAGACCC
 15 CATGCCTTAT TGCAAGGGCA GGGTCCAAT GTGACAGCCT AAGTTCTGC TCAGTGTACC
 AGAAGAATTG GATCACACGT GGGCTGGAAG GATGAGCACA AGGTTTTATT GAGTGGTGG
 GGTGGCTCTC CGCGAGACGA CTCTCAGCCA GAGAAGGAGG ATGGAGGGGA AAGGTGTTCT
 TCCCCGGAG TCCACTTGT CCTGTTCTCC TCTGGTTCA AATACCTCTT CCCTCCTTTT
 CTGCTGTGCT GTCCCACAC TCTCCACAGC TCTGTGCCGC TCTGTTCCCT TGCTCCTCTG
 GATGTTCAGC TCCTTGTATC TGTGCCTGCT AAGATTGG GTTTATATGG GGGAAAGGATG
 20 GGGGCATGG CGGGCCAAAA GGCACCTTT TTGGTGTGAA AACAGAAATG CCTGCTTCT
 CTTAGGGCCT TAGGTCTTCA GGCTTGAGGG TGGGGCCTTT GCTGAGGGAC CACCCCTTTC
 TACCCAGTAT TTCCCTGTCT CCTGTCCATA TCAACATTAC TAACTTTTC ATCTGCAGAC
 TAATAATGCT AAGGTGTGGC ATTTCACAC TGTGAGACTA TGTGAAGGTT TTTCTGTCC
 AACTGATGGC ATCCTCCCAT AATTCTACCC CTTTCTAAA AGAATTTTT GCAGTATTTC
 25 TCCAAGTTA TTCTAGAGAA TTTCTTGTCT GTGAAATGCT CTAGTTAAC CAAATTGGAA
 AACGGAGCAT ATCATATCCC CTTCTCAAAT TCACCAAAGT GAAGTCCTAA TGTGCTTAA
 TGTATCTGCA TGAGACAGGA AGCTGAGATC TATTCAACAA CAAAATCCA AACAAAGCATC
 AAGAGGAGGA GTGTTAGCAC TTGAGCCTAG GGAGACTGTG GCTCCTGCCT GAAAGATGGG
 AGCCCTCAGTC ACAGCTGCTT TACCAAGTGT CATATGCTAT GTTTCTGAGG ACTCCTGCTA
 30 AAGCTCCCTT CTCCCTCCAG CCAACCACCT TTGTTTTAGA CAAGGGCTGG GTTTATGAAG
 GACTGTTTTC ATGACTAAAG CTTTATAGAA GTTTAAGAT AAGGAGATGG AATTGAGTGA
 ACTAGGAAAT ATGAAAGCAG ATATTATAAT CTGGCTTCCT GATTTTCAC TAGCATTTC
 GTTTATAAAT TAGTTCTGTT CTAAGAATCC AATGACGTAA TAGAAACTCT CAAAGATTCT
 TAACTTGAGA TATAGGGAGT CTTGAAACT GCTGAAATTA CAGACAGCAT TTATGTTA
 35 TGAGCATTTC TGAATCTAGA GCTTCACTA GATTTGTAAA GAATGTGGC CAAAAGATTA
 AGAGCCAATC CTGTATCTTG TACTCAAAT GTTTGTAATT CTCACCTTTT TATCCCCCAG
 ACTCCTCTTG CTCTCTCTTC ATTTCACAG TGTTACTGGA AAGGGGTCT GATCCAGACC
 CCAAGAGAGG GCTCTGGAT CTTGTGCAAG AAAGAATTAA GGGCGAGTCC TCAATGCAA
 GTGAAAGCAA GTTTATTAAG AAAGTAAAGG AATAAAAGAA TGGCTACTCC ATAGACAGAG
 40 CAGCCCTGAG GGCAGCTGGT TGTCCATTAA TATGGTTATT TCTTGATTGT ATGCTAAACA
 AGGGATGGAT TATTCTAGCC TCCCCCTTTT AGACCATAGA GGGTAACCTC CTGACATTGT
 CATGGCATTG TTTTTTTTT TTGACAGAGT CTCACTGTGT TGTCCAGGCT GGAGTGCAGGT
 AAAGCAATCT TGGCTCACTG CAACCTCTGC CTCCTGGGTT TAAATGATTC TCCTGCCTCA
 GCCTCCCAAG TAGCTGGGAT TACAGGCATG ACCCACCACG CCCAGCCAAT TTTTGTATTT

TTAGTTGAAA CAGGGTTTTG CCATGTTGGC CAGGCTGGTC TTGAACTCCT GACTTCAGGT
 GATCTGCCTA CCTTGGCCTC CCAAAGTGCT GCAATTACAG ATGTGAGCCA CCAAACCTGG
 CCTGTATGG CATTCTTAAA CTGTCATGGT GCTGGTGGGA GTGTAGCAGT GAGGAAGACC
 AGAGGTCACT CTCATCGCCA TCTTGGTTTT GGTGGGTTTT AGCTGGCTTC TTTACTGCAG
 5 CCTGTCTGTT CTATCAGCAA GGTCTTTATG ACCTGAATCT TGTGCTGACC TCCTATCTCA
 TTCTATGACT TAGAATGCCT TAACTGTCTG GGAATGCAGC CCAGTAGGTC TCAGCCTCAT
 CTTACCCAGC TCCTATTCAA GATGGAGTTG CTCTGGTCA AACACCTTTG ACAACATCAT
 TAAGCCTCAG TTCTCACACT GTTTTGTTT TGTTTGTTG TGCAATGATG GGCAAATCTC
 10 TGCCCTATAG GATGGTAGAA AAAAGGAAC TAATATTGTA ATGACTGTTT TGTGCCAGAT
 AATTGCTTTA AACTGTATCA GCATCTTATT TAGTCCTGTT AATGATATGA ATGTTATCTT
 CATAACAGCT GCCATTTAT TAAGGACTTA TCAGAGAAAA ACACTGTTCT AAGCACTTGT
 TACCCATTAT TACATTGAAT TTTCATATAAC ACCCTTGAG GTAAGCATGA TTATACCCAC
 15 TTTAATAGAA GTGAACTGTA GTTTGTAAT GTTAGGTTCC TTGCCAAATG TTACACAGAT
 AGTAAGTGT AAAATCATAT GCCCTGAAAT TACATTATGC TGCCAAACTT AAATTTCTTT
 TTTATCCTTT ATATTAGTAT ATTCTTAGGT TTAAACAAAGA CAACTAGTTA ACACATACTA
 GATTTGTCC ACAGTTCCCTG GCTCATAACT CCCATAGCCC TTGTCACTAT CTTTAANCG
 TTGGGGCACA TTAGGCCTCA GAAGTAGGCC TCANGAAAAC AGAGTCTCTC TCTCTCTCTC
 TCTCTGATCT TCTTCCGCC CTCCTCTCACC TGCCCAGGGC AGCACTTTAA TCTTCTCCTG
 CCTTTCTGAT CTTGGGTCA AAGACCTTCA TTTCCAAAGA TGTCTGTGT CATACCTAA
 20 AGGAAGGAAC ACTGAACAGA GAGAGGCTCA GAAGAACATG GACAGGCCTT GCTGTGTTA
 CATCATCCCC TTTATGTCCA GTCACATCTC TACATGGTT TCAGTTGTGC CTATTGATG
 AAGTCCCCAT ATAAGGCTCA CAAGGACAGG GTGCAGAGAG CTTCCAGATA GCTGAACAAG
 TGGAAAGTTCC TGGAGGGTGG CGTGTTCAGG GAGGGCATGG AAGCTGTGTG CCCCTTCCCC
 CATACCTTGC CCTACTCATT TCTTCATCTG TTTCATTTGT AGTATCTTT ATAATAAAC
 25 ACTAAACATT AGTTAGTATT TCTCTGAGTT CTGTGAGTC CTCTAGCAA TTAATTGAAC
 CCAAGGAGGG TGTCTAGGA TCCCCNACAT TATAGCTGGT TGGCCAGAAAG CACAGGTAAA
 CAACCTAGGG CTTTCAATTG GCATGAGAAG TAGGGGGCAG TTTTGTGGGA CGGAGCCCTC
 AGCCTGTGAG ATCTGATGCC ATCTCTAAGT ACACAGTGTG AAAACTGGAT TGGAGGACAC
 CCAGCTAGTA TTCACTGTGA AATTGGTTGC TTGCTTGATT TGTGGGGAAA AACCCACATG
 30 CATTGATCA CAGAAGTCTT TTGTGTTGAC AGTTGATAGT GTTCAGTGAG AGAATTAAAA
 AAAAATTGAG TTTCTCTTC AACATACTCT CTCAATGTGA AACACAGAA ACTATTCCA
 TTCAAAGATG GAAATGGTTT GTTGCATCT TAGTTTTAT TTATACATCT TAGAAGAAAT
 GTCCAAGCTT GTTTTTTCT CTCACCCCTAT ATATAAAATT ACCTATGAGG CACAGATTT
 TATGATCCTT GATTATATAG ACTTTGTCCA AATTGTGTGT TTTATAGCAT TACTGTAACT
 35 TGTTATAGTA ATCTTTGTGT ATATTATGTC TCTTAACATT GTCTTCCATA TTGTTAATGA
 CCATCTCATA TTTATCTCTG TATCATGTAT ATCTTCACCC AATGTGACTG GCTTAGGAGA
 AAAAATTAGT GAACAATTAA CTAGTGGTTG TGTAATCTAT ACAATTGTCA AGGTTACAAT
 TGCTATTTTT GAAGAAATCG TTGTTGGTTT TCTCTTGTT TCATCTCAGT TCCATTGTT
 CAAGGATTCC TTTTTTTTTT TTTTTTTTTT TTTTGAGGC GGAGTCTTGC TCTGTCACCC
 40 GGGCTGGAGT GCAGTGGTGC AATCTCGGCT CGCTGCAAGC TCCACCTCCT GGGTTCATGC
 CATTCTCCTG CCTCAGCCTC CCGAGTAGCT GGGACTACAG GCACCCGCCA CCACGCCAG
 CTAATTCCCC TTTGTATTT TAGTAGGGAC GGGGTTTCAC TGTGTTAGCC AGGATGGTCT
 CAATCTCCTG ACCTCGTGT ACCTCGTGT CGGCCCGCCT CGGCCTCCCA ATGCTGGGAT TACAGGCGTG
 AGCCACCGCG CGTGGCCCT TGTCTAGTAT TCTTAACCTG TGTTATATCT AAGAGAAGGT

GTGAGAGGCG GGACCATTG TGGATAGGTG CAGAGGGCAT GAAGTAGCCC AGAAAGAAC
 TTTGCCATTG ACTAAATTAA AGCCTAGTAA AAAACATGTG GTAACCCTG AAAAACTAGA
 ACAGTGTGAT ATGACAATGC CCAGTTGAAT AAGAAATTCA AATAGTTTAT AGTAACAAAA
 AATAATTTT ACCACATTGT GACTAGTGCC CTAAAAACAA ATTTGAAGGA GCAGAGAGAG
 5 AGAGGAAGTG AGTTTCGCT GGGGGCTGGT GAGTGGAGGC TACTTTATGA GCAGTTTGA
 AAATTACAAG TTAGGGAAA ACTCTTAGGG GAGTTTAAG TGGTGAGCAA CAGGCACCTG
 AGGTTACAGG AGCAGCAGCA TAAAGATATA GAACAGAGAG GCCATTGGC AGTCTGGGG
 AAACTCCAAG AAGTCGAATG TGTCTAAAGC TGTGGTTCTC TCTCTGTGTG TGTGTGTGTG
 TGTGTGTGTG CGCGCGTGTG TTTGTGTGTG TGTGCATGTG TATGTATGTG TGAAATCGCT
 10 AATACAGAGA AGCTGTTAGA ACAATTTCCT CACTTCCCAA AAATGTCCGT TCATTCACTGG
 GCCTAGGCCTA CTCTCCTTTG TGTGTCTTAC AGCCCAACAC TGTCAATATAA GGTGTGGTTC
 TATAATGAAA CAGTTTATC TGTTGTTAC GGAGCTAGTC AGCCTGTGTT ATGCTTGTGTTA
 CTGGTTGAGG GTGTCCAGGT TCTTGGCTTC TTGAACAAAG AATTGGACAA AACTCACAAA
 TGAGGCAAGG AAAGAATGAA GCAACAAAG CGGAGATTAA TTGAAAATGA AAGCACACTC
 15 CACAGGATGG GAGTGGGCCT AAGTAAATGA TTCAAGGCC TGTTTACAGA ATTTTCTGGG
 GTTTCAATAC CGTCTAGAGG TTTCCATTG GTTACCTGGT GTATGCCTTA TGAAATGAA
 GAGGGTGGAG TTAAGTTACA AAGTCATTAA CTCAGTATAG GCCTGTGTT AATGGAGAGG
 GTGTTACTCC TGGGGGTTGT GGCCCAGTAA AACGGAGAGG ATGAAGTGAA GTGACAAAGC
 CCTTCGCATT CCTGCCATTG CTGAAGTGTG TCCACTTAT TTAGTTCTAG GAAGTCAGTG
 20 TGAATTGGCC TTATGTTCCC TGCCTCCAGA ACCTGTTCTC CTGCCTCACA TAGACCCTT
 TTTCCCTTG CCAACTTCAC TTCTTTACA GCCACCTCAC TACCAATGTG TCTGTCTCCT
 AAGTCAAACA TCAGTGATTC CTCTGTTTC TCTAAACCCCT TCTTATGTCT TCTACTTCTC
 ATCTCTTCT TGTTCAAAAAA TCTTCAAAA ATGAGTAAGA ATGCAGCTAT TCAGGCAAAC
 TAAAAATAAC ATCACAGTGA TATACAAAC CAGTGTCAATT TCACAAAGGA AAATTATCAA
 25 TACTAGATCC TGAAAAGAA ACAGCGAATG AAAGCCATT ACACAACCTCC ATTGTGTAAT
 TGACACATTG AATCACTCAT AAAACAGGTG CTCTGGTCT GAATCTAGAT CCTAGCTAGT
 CTGGTAGCTG AAATCATAGA ATTATAGTAG AGTTTAGGAA ATCATCCTCA AAGGAAAGAT
 TATATGTTGA TATCAAATGT ATATTCCTT TCTAGGCCA AACCTGCACC ACTGACTGCT
 GAAATACAGC AAAAGATTT GCATTTGCCA ACATCTGGG ACTGGAGAAA TGTTCATGGT
 30 ATCAATTTCG TCAAGTCCTGT TCGAAACCAA GGTAAAAAAA TAAGCCTAAG TTTTTGTTA
 ATTTGTTGG AACTATTAT TGAACAGTTG CTCTGTGTGA TGGATTTCGG GGATACCTAG
 ATGGAATGGG CATGATCCCC TTTTACAGA AATAGAAAAT AGGTGGCCTA TGAATTATTC
 TTTCCTTTA TATCCATGAC AAACTTAGT AAAAAATTCTT CTTTCTACT GAGCTTAGCA
 TTTATTCAGT ACTCTTCTCA ATATATTTTC CAGGTAGCTA GTGACAAATCA GAGTGATATG
 35 TAAGACAAAC TCATTTGTCC TCCTAGTAGG AAAGATTCTA GTAGAAGCAA AGAATTGTGT
 ACCATTCTGC AAGTGGTTTG TTGGAATCTT TCTTTGATAC CTGTTCCCTGT ATTCCCCCAC
 CCCCCATTAAT TTAGTCATTA ATTACTACAT GAACCCGTA AATAAAATCCT TAAATTATTT
 TCCTGGTAGA TTTTTGAGC TGTGTAAGGA CCTTTCAATT CACTTTACAT TAGAATACGA
 TATGTGATGG TAAGTATTAA CCCAGCTTCC TGAGTGATGG CGTGAGGGCG GAACCCAGGT
 40 TCATGAAAAG TATTCCTATT ATTGAATTAA ACCAGTTATT TTCAGAAGTT TGATACAATG
 TGAGTTGATT TCATAACATG TCCTCTAACT GCACCTGATA GCAAATTATC TTGTTATCCT
 GAGTTGTAGC CAACAATGAC TTGGAGAAC TATGCAATAC TCAGTTTAT TACTTCTAAG
 CTCATTTGA AGATAATACT ACCCATGGCT GATTGTTAC TATAAAATAG GTTTAGTATT
 TGCTGTCTGG AAACATTCTG ATTAGTGTCT CCTGGGAGGA TTATAAAATTT ATAGTACCCA

AAGANTAAAC NTGTTGTTTC CCTTTCCTAA ACTTTTAGTG NATAATNCAG TCTCTGCCGT
5 GTCTCATTTT CATCACTTGC CCTCCANAGC TCCCATCTCA CTGAATTCTT GCAGTGTCT
GAATGTTGAG AGCCCCAANG TGGGTCTTAT AACAGCCAGT CAGCAACATT TCTGTTTTC
ATCTGACACC AAGGGTCTCG TCTCTTGCT TTTCTACCAG TTATTCTGGG CTCTTCAGCT
CTAAAGAAAG TATAGGTCTT GAAATCTTTC CCTACCTTCT CAATTCCTG GGGAGGGCTT
10 CTTTGGAAAG TGGGATTGGA AATAAGATAA ATTTGAAGAT AATTGAGAAA TGAATGGAAA
GTGAAATTGA AGGGTCCATG TTAAGAGATT GCAAGTTATG CTATCACCAA ATAGATTTT
TGTGCCTGAG AGGATTAATT CATACTGCAT ATTATGTGTT GACTTTATCA TTGAGGTCTT
GGCACATGAT AGCATTGGCA TGATATAATT TGAGCTACTG ATACTATAGT GTTGCTTCTG
15 GTGTTGTTAC CAGAACATCT AAAATATATT AGGATTTTT TAATGGCAGA GGAAATGAAG
ACTAATATGA CATACTCCCT GTCCTAGTGA TTTACAGTTT AGCAAGACAT ACAAGCAAAA
CATTAAAGTA AAGCATGATA ACTACTTAA TAAAGCATT TTTAAATTCA TTGGTAACAT
AAGAGAAGGT AGAAGAGTTT AGCAAACCCCT TCCCAAAAGA AATGATTGAC AAGTTATATG
AGATAATAAT TCAGGGAAAG GAAATTGAC TTTCTAAAGC CAAATTATTT GACATTGGTT
20 TTCATATAGC TTGGTAAAAG CTGTTATTTT CTCCCAGTGT CTTTATTCTT TGACTGTTAT
AAGTATGATT TGTACAGAGA AATGGCAATT TCAAAACAGA GGGCTTTGAT GGATTAATTG
CTTGAAATTG ATCCCTCATC TACAGTATCT TGTCAGGTAC TTGAGAAAAAT AATGTAACAT
AAGTTTCCTC TTTGACTTT CTTTGGTAT TCTATACTGT AAGTTGGGG AAAAAGTATT
TTCTCTTCCT GCTAATTGGG CTACTTGAAA ATTCCCACCA ACTTTGCCAA TACCAGTGT
25 GTCACATTAA TTAGGAAATA CAGAGTGTTC TCATAATGCG TAAAGTCTAG TCCATTAAAG
TCTTGTTCAGA AAATGCTATT TATATTATTT GAGCAGGAAG GCAGAGACCT TAAACTGCTC
CACCCAATTC ATTTTACACA AAAGATTAAA AAGAAAAAAA CAGTGTCAA AGGTCAAGTG
CCCAGTGTG CACAACATAAT GATAGGCAAA ACAAGAGAGG AGGGGAAAAA AAAGAACCT
GACTCCGAC TCCTAGTTCA GTGTGTTCTT CACTGTAGA AGGTGCTGCT GAACATAGTT
30 ATACCATATG AAGATCACAC TATCTATTG AGATTGTAGA AATTTGATTA CTGCAGAGCT
CTGGCTGGGC TAGATTCACT TCTTATTCTT TATTGCATTG CAGTATTCTT AAGAGATAAA
TGGCTCTTT AGAACAGAT TGGCCTGGC TGTTGAAATG AGGCATTAAT TACCTTGGTA
GCTGACACAT TCTTACAGGT CAGGGGCTTG ATGAAGTTA TCTTCTTCCC TTGTTCTGT
GATTGCTCTG TAAATAGAAC ACATTCAAGAG CCCTTGAAATG CACTTGCAGC CTGTGCCTCC
35 CACAGTGATC GATGGTCAGA TAATGGGAGT TTAATGACCA GTACTGAGAG AGATTATTC
CATGGCTGCT ATGGGCCAAG AAAGCTGGGT GGTCAGAAAG GGACCTTTTC CAGACTCTCC
TGGGTTGTGT TATTCTTTG ACATTGGTTT CCTTTCATGA GCGGCCAAGG TGTATTTGTA
AGTTGTCCAG TGTTGCACAG CTAATGAGGG GCAAAACAAG AAAGGAGGAA AAAAAGTA
TCCTCACTCC TAGTTCACTG TGTTCTTCAC TGTTGGAAGG TACTGCTGAA CACAGTTATA
40 TGTTCACTG TGTTCACTG TGTTCTTCAC TGTTGGAAGG TACTGCTGAA CACAGTTATA
TGTTCACTG TGTTCACTG TGTTCTTCAC TGTTGGAAGG TACTGCTGAA CACAGTTATA
GGATAGTATC ATCTGCATGG TTTTGAAAC AAAAGATATA AGCTAATTTC GCCCTGTCTA
GTGACTACGA GACAGGGAGA GAAAATCTGA ATATTGTTA AAGTAGACAC AGACCCATAA
ATTGAAAAGG ACACAAATCC TGCCTTAGGA GACAGTAAGG CACTTGTCCC TGCTATCTAT
TAGCTGTGTG TCCTTGGACA GCTCATTGCT TCTTCTGAG CCACCGTTAC ACGTTATCTA
34

TCAAGTGAGC AGGTTGTACA CTAGATGAAT TCACAGGTCC TTCTTCCAAG TGCTTTCTA
ATCTTCATAA TTTAGATAAT CTCTCAGTAG CAAAACAGTG TACAATATGG TCAATCTGAG
ATTTTAGGG GGAGAATTTC AGGAAACATC AGAAATGGCA GTAGTTAAAA GGAAATAGGA
CTCACAGGCT GACTTCTCTC TATAACTTCA CATGGTAGAA GGGATGAGAG TTCTCTCTGG
5 AGCCTTTGT ATTAGGTACAC TAATCCAAA GGCCCCACCT CCTAAGACCA TCACTTTGG
GATTGGGGTT TTAACATGAT TTTAGAGGGA CATAAACCTT CATCTCATTG CATTAGAGAT
ATTTGAAAGC TCAGCTCACG TGTATTCTC CACAGCTCAC ATGTATTCCCT CCATCACCCA
ACCTGATGGC TTTGAACGTT GTAGACATAA ATCCTTCAT CATTATCAAG AATATTGCCA
AAAGCTTCTC AGAAATTATG AGGGGTTTT TTAGTTCTA AAATATTCTC AAAGAAAGTC
10 CCATGTACTA ATGTTTGCTT TTTGATGAAA AAGGATGAAA TCTTAATGAT TGCCTTAATA
AGCTCAACAA TGCTTGTAG TTGAGTCTTC TTATTGTGCT GATTCTTATA ACAACACA
TTCAGTATAA ACATTAATGT ATGTGATTCA CTAAGGTTTT TGCAATGATT TCTGTGAGGC
TTCAGATGTC TCTTGGATTA TGTGTCTTT TTTCATCGCC AGCATCCTGT GGCAGCTGCT
ACTCATTGTC TTCTATGGGT ATGCTAGAAG CGAGAACTCG TATACTAAC AACAAATTCTC
15 AGACCCCAAT CCTAAGCCCT CAGGAGGTTG TGTCTTGTAG CCAGTATGCT CAAGGTAAGT
GTTGCATTTC AGACACCATT TATGAGCTAT TTACCTGTGT GCAGCTGGCT GTTGGTGGCA
AAGGCAAAG GATGATGCAG TAGAGAGAGC GCAGTGTCTA TAGTCAGAAA ATCTGAGTGC
AAGTCTGGCC CTATCACTTA TTAATGGATG ATTGCTCATG GAATTTACTG TACCATCCAG
CAAATGTCA ATAGTTACTA TATATTGAGT GAGCTCTGCT TGTACATAT ATTGCCTAAC
20 AATGCTAAA ACTCTGAGAA GTAGTAAGTA TAATCCCTAT TTATGGCGG GGAACAGGAA
CTAAGAAATT TTTCTAATAA TTGAGGTC TCACAGCTTT TAGCATTGGA GTTCACTTC
TAATCATCGT CTCCAAAACC CAACTTTAT TAAAACATA CTAACACTGG TTTCTCTCG
GGAGAATTTC AAAATTCTGT ACTTAGGGCT GGGCACAGTG GCTTATGCCT ATAACCCCTAT
CACTTGGGA GGCTGAGATG GGTGAATCTC TTGAGTCCTA GAGTTGAGA CCAGCCTGGG
25 CAACACGGCG AAACCCCTTC TCTATTAAAA ATACAAAAAA TTAGCTGGC GTGGTGGTGT
GTGCTTGTAG TCCCAGCTAT TCAGGAGGCG GAGGTGTAAG AATCACCTGA GCCCAGGAGG
TCAAGGCTGC AGTGAGCCGA TATCATGCCA CTGCACTCCA GCCTGGCAA ACGGAGTGAG
GCCCTGTTAT GAAAAAAAAAA AAATCTGTAC TTAGGCTTT AGATCAGGCT GTATGTGATG
TATGTCAAA ACACAGCTAT AATTGATTGA GGGAGAAACG TTACCATTTT AAAGTTTATG
30 CTTTCAAGCC CAGATTGGC CACTAGGAAT TTCCCAGCTC ACTAGTGAAA CTGCTGATGA
GTGATTATTT GCCAGTGAGC CTTTCATTCT TTCTAAAATA TGTACTACTA GTTGTGACTT
GTAGGCTATAA GGGGCTATAA TATATCAAGA CAATCTTTAT CCTCATGAAG CTTACAGTTA
AGTAAGAGAT AGAGATTAAGA TAATTATAAC AACAGAGTGA AGAACAGTGA AGAAAAAGTA
CAGAGTTATA ATATATATAA TAGGGCCAGG ACTGCATGAG GAAGGTAGGA AAGACATTTC
35 GGCAAGAGGT TGTCAGGGAA AAGACTTGCT TGAGAAAGAG CCAAGTTGTG GGGTCTGGCT
GCTTAGCAAT GACCATAATA CCTAACTTTT GCTATTTTA CATGAAGTAA CTAATTTAAC
CCTATGAGGA AAGTACTACT ACCATCTAGA TTTTACAGGT AAGTAAGCAG AGATACAGAG
AAGTTAAACT CTTCACACGG CTTTGGCTTT AAACCTATAT AGGCTTCAGA GCCTCCCCAC
TTAACCACTT TGCCATAGCT ACATCCATAT TAGGTGCTAA GTAGATATCT GTTAAGTADA
40 AGGACGGATGA AAGGATAGTT AGCTAGTTGG AAAATGGATG GATGAATGAA GTGATGCTTA
AGCTAAGAAC AACTTTCAAGG GGTAACATGC AAAGAATAAT GGAGCAAAGA AGAAAAAAATA
GAAAATGGGA TAATCCTTTT TCTACTAAGG GGTAACCAGT GGTGTTATTG ATCTTCAGGC
TGTGAAGGCG GCTTCCCAT CTTTATTGCA GGAAAGTACG CCCAAGATTT TGGGCTGGTG
GAAGAAGCTT GCTTCCCTA CACAGGCACT GATTCTCCAT GCAAAATGAA GGAAAGACTGC

TTTCGTTATT ACTCCTCTGA GTACCACTAT GTAGGAGGTT TCTATGGAGG CTGCAATGAA
 GCCCTGATGA AGCTTGAGTT GGTCCATCAT GGGCCCATGG CAGTTGCTTT TGAAGTATAT
 GATGACTTCC TCCACTACAA AAAGGGGATC TACCACCACA CTGGTCTAAG AGACCCTTC
 AACCCCTTG AGCTGACTAA TCATGCTGTT CTGCTTGCG GCTATGGCAC TGACTCAGCC
 5 TCTGGGATGG ATTACTGGAT TGTTAAAAAC AGCTGGGCA CCGGCTGGGG TGAGAATGGC
 TACTTCCGGA TCCGCAGAGG AACTGATGAG TGTGCAATTG AGAGCATAGC AGTGGCAGCC
 ACACCAATTG CTAATTGTA GGGTATGCCT TCCAGTATTT CATAATGATC TGCATCAGTT
 GTAAAGGGGA ATTGGTATAT TCACAGACTG TAGACTTTCA GCAGCAATCT CAGAAGCTTA
 10 CAAATAGATT TCCATGAAGA TATTGTCTT CAGAATTAAA ACTGCCCTTA ATTTTAATAT
 ACCTTCAAT CGGCCACTGG CCATTTTTT CTAAGTATTC AATTAAGTGG GAATTTCCTG
 GAAGATGGTC AGCTATGAAG TAATAGAGTT TGCTTAATCA TTTGTAATTC AAACATGCTA
 TATTTTTAA AATCAATGTG AAAACATAGA CTTATTTTA AATTGTACCA ATCACAAAGAA
 AATAATGGCA ATAATTATCA AAACCTTTAA AATAGATGCT CATATTTTA AAATAAAGTT
 TTAAAAA
 15

The corresponding cDNA sequence for CTSC is provided below (SEQ ID NO: 2):

1 aattcttcac ctcttttctc agctccctgc agcatgggtg ctggggccctc cttgctgctc
 20 61 gcccctcc tgctgcttct ctccggcgac ggccgcgtgc gctgcgacac acctgccaac
 121 tgcacctatc ttgacctgct gggcacctgg gtcttccagg tgggctccag cggttcccag
 181 cgcgtatgtca actgctcggt tatgggacca caagaaaaaa aagttagtggt gtaccttcag
 241 aagctggata cagcatatga tgaccttggc aattctggcc atttcaccat cattacaac
 301 caaggcttg agattgtgtt gaatgactac aagtgggttgc cttttttaa gtataaagaa
 361 gagggcagca aggtgaccac ttactgcaac gagacaatga ctgggtgggt gcatgatgtg
 421 ttggccgga actgggcttg tttcaccggaa aagaagggtgg gaaactgcctc tgagaatgtg
 481 tatgtcaaca cagcacaccc taagaattct caggaaaagt attctaataag gctctacaag
 541 tatgatcaca actttgtgaa agctatcaat gccatcaga agtcttggac tgcaactaca
 601 tacatggaat atgagactct taccctgggaa gatatgatta ggagaagggtgg tggccacagt
 661 cgaaaaatcc caaggcccaa acctgcacca ctgactgctg aaatacagca aaagattttg
 721 catttgc当地 catcttggaa ctggagaaat gttcatggta tcaattttgt cagtcctgtt
 781 cgaaccaag catcctgtgg cagctgctac tcatttgctt ctatgggtat gctagaagcgc
 841 agaatccgta tactaaccaa caattctcag accccaaatcc taagccctca ggagggtgtg
 901 tcttgttagcc agtatgctca aggctgtgaa ggccgcttcc cataccttat tgcaggaaag
 961 tacgccccaaatggct ggtggaaagaa gcttgccttcc cttacacagg cactgattct
 1021 ccatgcaaaa tgaaggaaga ctgctttcgat tattactcct ctgagtgatcca ctatgttagga
 1081 ggtttctatg gaggctgcaa tgaagccctg atgaagcttg agttgggtcca tcatggcc
 1141 atggcagttg ctttgaagt atatgatgac ttccctccact aaaaaaaggg gatctaccac
 1201 cacactggtc taagagaccc ttcaaccccc ttgagctga ctaatcatgc tttctgtt
 1261 gtgggctatg gcactgactc agcctctggg atggattact ggattgttaa aaacagctgg
 1321 ggcacccggct ggggtgagaa tggctacttc cggatccgca gaggaactga tgagtgtgca
 1381 attgagagca tagcagtggc agccacacca attctaaaat ttttaggtat gccttccagg
 1441 atttcataat gatctgcatac agttgttaaag gggaaattgggtt atattcacag actgttagact

1501 ttcagcagca atctcagaag cttacaaata gattccatg aagatatttg tcttcagaat
1561 taaaactgcc cttaatTTTA atataccTTT caatcgGCCa ctggccATTt ttttctaagt
1621 attcaattaa gtggGAATTt tctggAAGAT ggtcagCTAT gaagtaatAG agtttgcTTA
1681 atcatttGTA attcaaACAT gCTATAATTt ttAAAATCAA tgtgAAAACA tagacttATT
5 1741 ttAAATTGT accaatcaca agAAAATAAT ggcaATAATT atcaAAACTT ttAAAATAGA
1801 tgctcatATT ttAAAATAA agTTTTAAA ataactgc

The wild type CTSC protein sequence is set forth below as SEQ ID NO: 3:

MGAGPSLLLALLLSDGAVRCDPANCTYLDLLGTWVFQVG
SSGSQRDVNCVMPQEKKVVVLQKLDTAYDDLGNSGHFTI IYNQGFEIVLNDYKWF
AFFKYKEEGSKVTTYCNETMTGWVHDVLGRNWACFTGKKVGTASENVVNTAHLKNSQ
EKYSNRLYKYDHNFVKAINAIQKSWTTATYMEYETLTLCMDIIRRSGGHRSRKIPRKPA
PLTAEIQQKILHLPTSWDWRNVHGINFVSPVRNQASCGSCYSFASMGMLEARIRILTN
NSQTPILSPQEVVSCSQYAQGCEGGFPYLIAKGYAQDFGLVEEACFPYTGTDSPCKMK
EDCFRRYYSEYHYVGGFYGGCNEALMKELVHGPMAVAFEVYDDFLHYKKGIYHTG
LRDPFNPFELTNHAVLLVGYGTDASGMDYWIVKNSWGTGWGENGYFRIRRGTDCAI
ESIAVAATPIPKL.

Papillon Lefevre syndrome (PLS) is an autosomal recessive disorder characterized by palmoplantar hyperkeratosis and severe early onset periodontitis that results in the premature loss of the primary and secondary dentitions. The 46 kb *CTSC* gene consists of 7 exons and is mutated in PLS patients. Sequence analysis of *CTSC* from PLS affected individuals from thirty-two Turkish families identified four different mutations. An exon 6 nonsense mutation (856C->T) introduces a premature stop codon at amino acid 286. Three exon 2 mutations were identified including a single nucleotide deletion (1047delA) of codon 349 introducing a frameshift and premature termination codon, a two base pair deletion (1028-1029delCT) that results in introduction of a stop codon at amino acid 343, and a G->A substitution in codon 429 (1286G->A) introducing a premature termination codon. All PLS affected individuals examined were homozygous for cathepsin C mutations inherited from a common ancestor. Parents and siblings heterozygous for cathepsin C mutations do not show either the palmoplantar hyperkeratosis or severe early onset periodontitis characteristic of PLS. In

addition to the 5 families described above, Table I summarizes CTSC mutations identified in 27 other families presenting with symptoms of PLS.

Haim-Munk syndrome is a rare condition associated with congenital palmoplantar keratosis, pes planus, onychogyrphosis, periodontosis, arachnodactyly and acroosteolysis. In an additional embodiment of the invention, a mutation in cathepsin C causing Haim-Munk Syndrome has been identified. It appears that substitution of an A for a G at CTSC nucleotide position 857 in Exon 6 is responsible for this syndrome in patients.

Based on the data presented herein, it appears that additional mutations or functional polymorphisms are associated with other pathological conditions, including, but not limited to prepubertal periodontitis (PPP), early onset periodontal disease or other forms of gum disease. For example, as shown herein, PPP is caused a substitution of a G for an A at position 1040 in the CTSC coding sequence. Thus, the invention also provides methods for screening the CTSC gene for alterations associated with these disease states.

I. Preparation of Altered Human CTSC-Encoding Nucleic Acid Molecules, CTSC Proteins, and Antibodies Thereto

A. Nucleic Acid Molecules

Nucleic acid molecules encoding the human CTSC proteins of the invention may be prepared by two general methods: (1) synthesis from appropriate nucleotide triphosphates, or (2) isolation from biological sources. Both methods utilize protocols well known in the art. The availability of nucleotide sequence information, such as a DNA having the sequence of SEQ ID NOS:1-2 enables preparation of an isolated nucleic acid molecule of the invention by oligonucleotide synthesis.

Synthetic oligonucleotides may be prepared by the

phosphoramidite method employed in the Applied Biosystems 38A DNA Synthesizer or similar devices. The resultant construct may be purified according to methods known in the art, such as high performance liquid chromatography (HPLC). Long, double-stranded polynucleotides, such as a DNA molecule of the present invention, must be synthesized in stages, due to the size limitations inherent in current oligonucleotide synthetic methods. Thus, for example, a 4.7 kb double-stranded molecule may be synthesized as several smaller segments of appropriate complementarity. Complementary segments thus produced may be annealed such that each segment possesses appropriate cohesive termini for attachment of an adjacent segment. Adjacent segments may be ligated by annealing cohesive termini in the presence of DNA ligase to construct an entire 4.7 kb double-stranded molecule. A synthetic DNA molecule so constructed may then be cloned and amplified in an appropriate vector.

Nucleic acid sequences encoding the altered human CTSC proteins of the invention may be isolated from appropriate biological sources using methods known in the art. In a preferred embodiment, a cDNA clone is isolated from a cDNA expression library of human origin. In an alternative embodiment, utilizing the sequence information provided by the cDNA sequence, human genomic clones encoding altered CTSC proteins may be isolated.

Table 1 sets forth several different mutations associated with particular PPKs and PPP. Altered CTSC-specific probes for identifying such sequences may be between 15 and 40 nucleotides in length. For probes longer than those shown above, the additional contiguous nucleotides are provided within SEQ ID NOS:1 and 2.

Additionally, cDNA or genomic clones having homology with human CTSC may be isolated from other species using oligonucleotide probes corresponding to predetermined sequences within the human CTSC encoding

nucleic acids.

In accordance with the present invention, nucleic acids having the appropriate level of sequence homology with the protein coding region of SEQ ID NO:1 may be identified by using hybridization and washing conditions of appropriate stringency. For example, hybridizations may be performed, according to the method of Sambrook et al., Molecular Cloning, Cold Spring Harbor Laboratory (1989), using a hybridization solution comprising: 5X SSC, 5X Denhardt's reagent, 1.0% SDS, 100 µg/ml denatured, fragmented salmon sperm DNA, 0.05% sodium pyrophosphate and up to 50% formamide. Hybridization is carried out at 37-42°C for at least six hours. Following hybridization, filters are washed as follows:

(1) 5 minutes at room temperature in 2X SSC and 1% SDS;
(2) 15 minutes at room temperature in 2X SSC and 0.1% SDS; (3) 30 minutes-1 hour at 37°C in 1X SSC and 1% SDS; (4) 2 hours at 42-65°C in 1X SSC and 1% SDS, changing the solution every 30 minutes.

Nucleic acids of the present invention may be maintained as DNA in any convenient cloning vector. In a preferred embodiment, clones are maintained in a plasmid cloning/expression vector, such as pBluescript (Stratagene, La Jolla, CA), which is propagated in a suitable *E. coli* host cell.

Altered CTSC-encoding nucleic acid molecules of the invention include cDNA, genomic DNA, RNA, and fragments thereof which may be single- or double-stranded. Thus, this invention provides oligonucleotides having sequences capable of hybridizing with at least one sequence of a nucleic acid molecule of the present invention, such as selected segments of the DNA having SEQ ID NO:1. Also contemplated in the scope of the present invention are oligonucleotide probes which specifically hybridize with the mutated CTSC genes of the invention while not hybridizing with the wild type sequence under high stringency conditions. Primers

capable of specifically amplifying the altered CTSC encoding nucleic acids described herein are also contemplated herein. As mentioned previously, such oligonucleotides are useful as probes and primers for detecting, isolating or amplifying altered CTSC genes.

5 Antisense nucleic acid molecules may be targeted to translation initiation sites and/or splice sites to inhibit the expression of the CTSC gene or production of the CTSC protein of the invention. Such antisense molecules are typically between 15 and 30 nucleotides in length and often span the translational start site of 10 CTSC encoding mRNA molecules.

15 Alternatively, antisense constructs may be generated which contain the entire CTSC cDNA in reverse orientation. Such antisense constructs are easily prepared by one of ordinary skill in the art.

It will be appreciated by persons skilled in the art that variants (e.g., allelic variants) of CTSC sequences exist in the human population, and must be taken into account when designing and/or utilizing 20 oligonucleotides of the invention. Accordingly, it is within the scope of the present invention to encompass such variants, with respect to the CTSC sequences disclosed herein or the oligonucleotides targeted to specific locations on the respective genes or RNA transcripts. Accordingly, the term "natural allelic 25 variants" is used herein to refer to various specific nucleotide sequences of the invention and variants thereof that would occur in a human population. The usage of different wobble codons and genetic polymorphisms which give rise to conservative or neutral 30 amino acid substitutions in the encoded protein are examples of such variants. Such variants would not demonstrate altered CTSC activity. Additionally, the term "substantially complementary" refers to 35 oligonucleotide sequences that may not be perfectly matched to a target sequence, but such mismatches do not

materially affect the ability of the oligonucleotide to hybridize with its target sequence under the conditions described.

5

B. Proteins

Full-length, altered, human CTSC proteins of the present invention may be prepared in a variety of ways, according to known methods. The proteins may be purified from appropriate sources, e.g., transformed bacterial or animal cultured cells or tissues, by immunoaffinity purification. However, this is not a preferred method due to the low amount of protein likely to be present in a given cell type at any time. The availability of nucleic acid molecules encoding CTSC protein enables production of the protein using *in vitro* expression methods known in the art. For example, a cDNA or gene may be cloned into an appropriate *in vitro* transcription vector, such as pSP64 or pSP65 for *in vitro* transcription, followed by cell-free translation in a suitable cell-free translation system, such as wheat germ or rabbit reticulocyte lysates. *In vitro* transcription and translation systems are commercially available, e.g., from Promega Biotech, Madison, Wisconsin or Gibco-BRL, Gaithersburg, Maryland.

Alternatively, according to a preferred embodiment, larger quantities of CTSC protein may be produced by expression in a suitable prokaryotic or eukaryotic system. For example, part or all of a DNA molecule, such as a DNA having SEQ ID NOS:1 or 2 containing an alteration set forth in Table 1 may be inserted into a plasmid vector adapted for expression in a bacterial cell, such as *E. coli*. Such vectors comprise the regulatory elements necessary for expression of the DNA in the host cell positioned in such a manner as to permit expression of the DNA in the host cell. Such regulatory elements required for expression include

promoter sequences, transcription initiation sequences and, optionally, enhancer sequences.

The human CTSC protein produced by gene expression in a recombinant procaryotic or eukaryotic system may be purified according to methods known in the art. In a preferred embodiment, a commercially available expression/secretion system can be used, whereby the recombinant protein is expressed and thereafter secreted from the host cell, and readily purified from the surrounding medium. If expression/secretion vectors are not used, an alternative approach involves purifying the recombinant protein by affinity separation, such as by immunological interaction with antibodies that bind specifically to the recombinant protein or nickel columns for isolation of recombinant proteins tagged with 6-8 histidine residues at their N-terminus or C-terminus. Alternative tags may comprise the FLAG epitope or the hemagglutinin epitope. Such methods are commonly used by skilled practitioners.

The human CTSC protein of the invention, prepared by the aforementioned methods, may be analyzed according to standard procedures. For example, such protein may be subjected to amino acid sequence analysis, according to known methods.

The present invention also provides antibodies capable of immunospecifically binding to proteins of the invention. Polyclonal antibodies directed toward altered human CTSC proteins may be prepared according to standard methods. In a preferred embodiment, monoclonal antibodies are prepared, which react immunospecifically with the various epitopes of the CTSC protein described herein. Monoclonal antibodies may be prepared according to general methods of Köhler and Milstein, following standard protocols. Polyclonal or monoclonal antibodies that immunospecifically interact with altered CTSC proteins can be utilized for identifying and purifying such proteins. For example, antibodies may be utilized

for affinity separation of proteins with which they immunospecifically interact. Antibodies may also be used to immunoprecipitate proteins from a sample containing a mixture of proteins and other biological molecules. Other uses of anti-CTSC antibodies are described below.

5
10
II. DETECTION OF KERATODERMAL DISORDERS/DYSPLASIAS and PERIODONTAL DISEASE-ASSOCIATED MUTATIONS AND DIAGNOSTIC SCREENING ASSAYS THEREFORE

Currently, the most direct method for mutational analysis is DNA sequencing, however it is also the most labor intensive and expensive. It is usually not practical to sequence all potentially relevant regions of every experimental sample. Instead some type of preliminary screening method is commonly used to identify and target for sequencing only those samples that contain mutations. Single stranded conformational polymorphism (SSCP) is a widely used screening method based on mobility differences between single-stranded wild type and mutant sequences on native polyacrylamide gels. Other methods are based on mobility differences in wild type/mutant heteroduplexes (compared to control homoduplexes) on native gels (heteroduplex analysis) or denaturing gels (denaturing gradient gel electrophoresis). Sample preparation is relatively easy in these assays, and conditions for electrophoresis required to generate the often subtle mobility differences that form the basis for identifying the targets that contain mutations are known to those of skill in the art. Another parameter to be considered is the size of the target region being screened. In general, SSCP is used to screen target regions no longer than about 200-300 bases.

Another type of screening technique currently in use is based on cleavage of unpaired bases in

5

10

15

20

25

30

35

heteroduplexes formed between wild type probes hybridized to experimental targets containing point mutations. The cleavage products are also analyzed by gel electrophoresis, as subfragments generated by cleavage of the probe at a mismatch generally differ significantly in size from full length, uncleaved probe and are easily detected with a standard gel system. Mismatch cleavage has been effected either chemically (osmium tetroxide, hydroxylamine) or with a less toxic, enzymatic alternative, using RNase A. The RNase A cleavage assay has also been used, although much less frequently, to screen for mutations in endogenous mRNA targets for detecting mutations in DNA targets amplified by PCR. A mutation detection rate of over 50% was reported for the original RNase screening method.

A newer method to detect mutations in DNA relies on DNA ligase which covalently joins two adjacent oligonucleotides which are hybridized on a complementary target nucleic acid. The mismatch must occur at the site of ligation. As with other methods that rely on oligonucleotides, salt concentration and temperature at hybridization are crucial. Another consideration is the amount of enzyme added relative to the DNA concentration. In summary, exemplary approaches for detecting alterations in CTSC encoding nucleic acids or polypeptides/proteins include:

a) comparing the sequence of nucleic acid in the sample with the wild-type CTSC nucleic acid sequence to determine whether the sample from the patient contains mutations; or

b) determining the presence, in a sample from a patient, of the polypeptide encoded by the CTSC gene and, if present, determining whether the polypeptide is full length, and/or is mutated, and/or is expressed at the normal level; or

c) using DNA restriction mapping to compare the restriction pattern produced when a restriction enzyme

cuts a sample of nucleic acid from the patient with the restriction pattern obtained from normal CTSC gene or from known mutations thereof; or,

5 d) using a specific binding member capable of binding to a CTSC nucleic acid sequence (either normal sequence or known mutated sequence), the specific binding member comprising nucleic acid hybridizable with the CTSC sequence, or substances comprising an antibody domain with specificity for a native or mutated CTSC nucleic acid sequence or the polypeptide encoded by it, the specific binding member being labeled so that binding of the specific binding member to its binding partner is detectable; or,

10 e) using PCR involving one or more primers based on normal or mutated CTSC gene sequence to screen for normal or mutant CTSC gene in a sample from a patient.

15 A "specific binding pair" comprises a specific binding member (sbm) and a binding partner (bp) which have a particular specificity for each other and which in normal conditions bind to each other in preference to other molecules. Examples of specific binding pairs are antigens and antibodies, ligands and receptors and complementary nucleotide sequences. The skilled person is aware of many other examples and they do not need to be listed here. Further, the term "specific binding pair" is also applicable where either or both of the specific binding member and the binding partner comprise a part of a large molecule. In embodiments in which the specific binding pair are nucleic acid sequences, they will be of a length to hybridize to each other under conditions of the assay, preferably greater than 10 nucleotides long, more preferably greater than 15 or 20 nucleotides long.

20 In most embodiments for screening for susceptibility alleles, the CTSC nucleic acid in the sample will initially be amplified, e.g. using PCR, to increase the amount of the analyte as compared to other
25
30
35

sequences present in the sample. This allows the target CTSC sequences to be detected with a high degree of sensitivity if they are present in the sample. This initial step may be avoided by using highly sensitive array techniques that are becoming increasingly important in the art.

The identification of the CTSC gene and its association with keratodermal disorders/dysplasias and periodontal diseases paves the way for aspects of the present invention to provide the use of materials and methods, such as are disclosed and discussed above, for establishing the presence or absence in a test sample of a variant form of the gene, in particular an allele or variant specifically associated with PLS, HMS or periodontal diseases. This may be for diagnosing a predisposition of an individual to PLS, HMS or periodontal disease. It may be for diagnosing PLS, HMS or periodontal disease in a patient with the disease as being associated with the altered CTSC gene.

This allows for planning of appropriate therapeutic and/or prophylactic measures, permitting stream-lining of diagnosis, treatment and outcome assessments. The approach further stream-lines treatment by targeting those patients most likely to benefit.

According to another aspect of the invention, methods of screening drugs for therapy to identify suitable drugs for restoring CTSC product functions are provided.

The CTSC polypeptide or fragment employed in drug screening assays may either be free in solution, such as gingival crevicular fluid, affixed to a solid support or within a cell. One method of drug screening utilizes eukaryotic or prokaryotic host cells which are stably transformed with recombinant polynucleotides expressing the polypeptide or fragment, preferably in competitive binding assays. Such cells, either in viable or fixed form, can be used for standard binding assays. One may

determine, for example, formation of complexes between a CTSC polypeptide or fragment and the agent being tested, or examine the degree to which the formation of a complex between a CTSC polypeptide or fragment and a known ligand is interfered with by the agent being tested.

Another technique for drug screening provides high throughput screening for compounds having suitable binding affinity to CTSC polypeptides and is described in detail in Geysen, PCT published application WO 84/03564, published on Sep. 13, 1984. Briefly stated, large numbers of different, small peptide test compounds are synthesized on a solid substrate, such as plastic pins or some other surface. The peptide test compounds are reacted with CTSC polypeptide and washed. Bound CTSC polypeptide is then detected by methods well known in the art.

A further technique for drug screening involves the use of host eukaryotic cell lines or cells (such as described above) which have a nonfunctional CTSC gene. These host cell lines or cells are defective at the CTSC polypeptide level. The host cell lines or cells are grown in the presence of drug compound. The rate of growth of the host cells is measured to determine if the compound is capable of regulating the growth of CTSC defective cells.

The goal of rational drug design is to produce structural analogs of biologically active polypeptides of interest or of small molecules with which they interact (e.g., agonists, antagonists, inhibitors) in order to fashion drugs which are, for example, more active or stable forms of the polypeptide, or which, e.g., enhance or interfere with the function of a polypeptide *in vivo*. See, e.g., Hodgson, (1991) Bio/Technology 9:19-21. In one approach, one first determines the three-dimensional structure of a protein of interest (e.g., CTSC polypeptide) or, for example,

of the CTSC-substrate complex, by x-ray crystallography, by nuclear magnetic resonance, by computer modeling or most typically, by a combination of approaches. Less often, useful information regarding the structure of a polypeptide may be gained by modeling based on the structure of homologous proteins. An example of rational drug design is the development of HIV protease inhibitors (Erickson et al., (1990) Science 249:527-533). In addition, peptides (e.g., CTSC polypeptide) may be analyzed by an alanine scan (Wells, 1991) Meth. Enzym. 202:390-411. In this technique, an amino acid residue is replaced by Ala, and its effect on the peptide's activity is determined. Each of the amino acid residues of the peptide is analyzed in this manner to determine the important regions of the peptide.

It is also possible to isolate a target-specific antibody, selected by a functional assay, and then to solve its crystal structure. In principle, this approach yields a pharmacore upon which subsequent drug design can be based. It is possible to bypass protein crystallography altogether by generating anti-idiotypic antibodies (anti-ids) to a functional, pharmacologically active antibody. As a mirror image of a mirror image, the binding site of the anti-ids would be expected to be an analog of the original molecule. The anti-id could then be used to identify and isolate peptides from banks of chemically or biologically produced banks of peptides. Selected peptides would then act as the pharmacore.

Thus, one may design drugs which have, e.g., improved CTSC polypeptide activity or stability or which act as inhibitors, agonists, antagonists, etc. of CTSC polypeptide activity. By virtue of the availability of cloned CTSC sequences, sufficient amounts of the CTSC polypeptide may be made available to perform such analytical studies as x-ray crystallography. In addition, the knowledge of the CTSC protein sequence

provided herein will guide those employing computer modeling techniques in place of, or in addition to x-ray crystallography.

5

III Therapeutics

A. Pharmaceuticals and Peptide Therapies

The discovery that mutations in the CTSC gene give rise to PLS, HMS, and periodontal disease facilitates the development of pharmaceutical compositions useful for treatment and diagnosis of these syndromes and conditions. These compositions may comprise, in addition to one of the above substances, a pharmaceutically acceptable excipient, carrier, buffer, stabilizer or other materials well known to those skilled in the art. Such materials should be non-toxic and should not interfere with the efficacy of the active ingredient. The precise nature of the carrier or other material may depend on the route of administration, e.g. oral, intravenous, cutaneous or subcutaneous, nasal, intramuscular, intraperitoneal routes.

Whether it is a polypeptide, antibody, peptide, nucleic acid molecule, small molecule or other pharmaceutically useful compound according to the present invention that is to be given to an individual, administration is preferably in a "prophylactically effective amount" or a "therapeutically effective amount" (as the case may be, although prophylaxis may be considered therapy), this being sufficient to show benefit to the individual.

B. Methods of Gene Therapy

As a further alternative, the nucleic acid encoding the authentic biologically active CTSC polypeptide could be used in a method of gene therapy, to treat a patient who is unable to synthesize the active "normal" polypeptide or unable to synthesize it at the normal

level, thereby providing the effect elicited by wild-type CTSC and suppressing the occurrence of "abnormal" CTSC associated with keratodermal disorders and dysplasias.

5 Vectors, such as viral vectors have been used in the prior art to introduce genes into a wide variety of different target cells. Typically the vectors are exposed to the target cells so that transformation can take place in a sufficient proportion of the cells to provide a useful therapeutic or prophylactic effect from the expression of the desired polypeptide. The 10 transfected nucleic acid may be permanently incorporated into the genome of each of the targeted cells, providing long lasting effect, or alternatively the treatment may 15 have to be repeated periodically.

A variety of vectors, both viral vectors and plasmid vectors are known in the art, see US Patent No. 5,252,479 and WO 93/07282. In particular, a number of viruses have been used as gene transfer vectors, including papovaviruses, such as SV40, vaccinia virus, herpes viruses including HSV and EBV, and retroviruses. 20 Many gene therapy protocols in the prior art have employed disabled murine retroviruses.

Gene transfer techniques which selectively target 25 the CTSC nucleic acid to oral tissues are preferred. Examples of this include receptor-mediated gene transfer, in which the nucleic acid is linked to a protein ligand via polylysine, with the ligand being specific for a receptor present on the surface of the 30 target cells.

The following methods are provided to facilitate the practice of the present invention.

35 **Family material and clinical diagnosis.**

Five Turkish families were described previously [8]. All available family members provided consent for

the study and were clinically examined. A diagnosis of PLS was made in individuals with severe early onset periodontitis and the clinical appearance of hyperkeratosis on the palmar and plantar surfaces. All affected individuals also had hyperkeratosis on the knees. DNA was isolated from peripheral blood samples from all available members from these nuclear families using standard techniques (Qiamp Blood Kit, Qiagen).

RNA Isolation, Amplification, and Tissue Expression Analysis.

Total RNA was generated from fresh tissue samples (gingiva, palm, sole, knee) using TRIZOL reagent (Molecular Research Center, Inc.; Cincinnati, OH) according to the manufacturer's protocol. To determine if cathepsin C was expressed in a given tissue, single-tube RT-PCR was carried out using the Access RT-PCR System (Promega; Madison, WI), following the manufacturer's protocol. A portion of each reaction was visualized following agarose gel electrophoresis in the presence of ethidium bromide. Amplification primers located within exon 6 **F** 5'-AGGAGGTTGTCTTGTAGCC-3' (nt. 857-877; SEQ ID NO: 4) and exon 7 **R** 5'-AGTGCCTGTGTAGGGGAAGC-3' (NT 981-962; SEQ ID NO: 5) produce an amplicon of 123 base pairs from cDNA. A standard PCR protocol was followed with an annealing temperature of 65°C.

GenBank accession numbers.

Full-length cDNA of CTSC (NM-001814) and full-length genomic DNA of CTSC contained within a BAC vector, Genbank accession number (AC011088). See SEQ ID NO: 1.

Cathepsin C Activity Assay

In unaffected non-carriers, cathepsin C activity ranges from 600-1200 µmol/min/ mg. As carriers of a

cathepsin C mutation do not have clinical manifestations, measurement of cathepsin C enzymatic activity can be used to determine whether at-risk family members are carriers. Enzymatic activity can also be used to determine if individuals marrying into a family are carriers. Carriers typically have approximately 50% of normal enzyme activity. Determination of enzymatic activity can also be used to establish a diagnosis of PLS when mutational studies of cathepsin C have been negative. This is important in assuring that a diagnosis of PLS has been properly given to an individual with clinical symptoms suggestive of PLS.

Viable leukocyte pellets are obtained from lithium heparinized whole blood by mixing blood with 3 volumes of 3% dextran in normal saline, and allowing the red cells to settle for 45 min at room temperature. Cells are pelleted by centrifugation at 1500 rpm for 5 min at 4°C. After washing and removal of contaminating red cells, leukocyte pellets are resuspended in dH₂O and sonicated on ice for 5 sec each for total of 6 blasts using a Sonic 300 Dismembrator. Protein concentration is determined by the Lowry method.

Enzymatic activity is determined by measuring hydrolysis of the synthetic substrate glycyl-L-arginine-7-amido-4-methylcoumarin at a final concentration of 5 mM using a modified method. All reactions are performed in duplicate. Twenty µl of leukocyte lysate are added to 200 µl of Na₃PO₄ buffer (0.1M, pH 6.5) in a 96 well plate and then substrate added. Reactions are allowed to proceed for 1 hr at room temperature at which time 10 µl of glycine-NaOH buffer (0.5M, pH 9.8) is added to stop the reaction. Fluorescence is determined using a Perkin-Elmer LS50B luminescence spectrometer at 370-nm excitation and 460-nm emission. The amount of NHMec released is determined by generating a standard curve using NHMec. Cathepsin C activity is reported as µmol

NH₂Mec released per min per mg protein.

Sequencing and mutation analysis.

PCR primers were designed to cover the entire cathepsin C gene in overlapping fragments, from 955 nucleotides 5' to the start codon to 240 nucleotides 3' to the termination codon using cathepsin C (DPP-I) sequence data (Accession # U79415; SEQ ID NO: 1). The PCR products were prepared for sequencing by excising the bands from the agarose gel and extracting the fragments using a Qiagen Gel Clean-up Kit. The sense and antisense strand of each PCR product were directly sequenced on an ABI Prism 310 Genetic Analyzer (Perkin-Elmer) using four dye terminator chemistry.

Approximately 1-3 ng of purified product and 3.2 pmol primer were added to premixed reagents from the ABI Prism Big Dye Terminator Cycle Sequencing Ready Reaction Kit, FS (Perkin-Elmer) and underwent a cycle sequencing reaction in a GeneAmp PCR System 9700 (Perkin Elmer).

The linear amplification started with a 10 s denaturation at 96°C, 5 s annealing at 50°C and 4 min extension at 60°C. The fluorescently labeled sequencing products were separated from residual reaction reagents using a Centri-Sep spin column (Princeton Separations, Aldelphia NJ) and electrophoresed on POP6 capillary at 1500 V for 30 min. Sequencing data were automatically collected and analyzed by the ABI Prism 310 software.

Table A:

**Primers Used to Determine Genomic Organization of
Cathepsin C**

Region	Primer Sequences
Intron 1	F: 5'- TGTCAACTGCTCGGTTATGGGTAA-3' (#6) R: 5'- TCGAGCTTCTCTTCGTACACCACT-3' (#7)
Intron 2	F: 5'- TGACTACAAGTGGTTGCCTTTT-3' (#8) R: 5' TGCTGCCCTCTCTTATACTGC-3' (#9)
Intron 3	F: 5'- GCCTCTGAGAATGTGTATGTCAAC-3' (#10) R: 5' CCTGCCCAAAATGAGATA-3" (#11)

Intron 4	F: 5' TCGAAAAATCCCAAGGTAATC-3' (#12) R: 5' GGGCCTAGAAAGGAATATACATT-3' (#13)
Intron 5	F: 5' AATTGTTCTGGAACTATTATTGA-3' (#14) R: 5' TCGCTTCTAGCATACCCATA-3' (#15)

Exon	Primer Sequences
1	F: 5'-GGCGATCAGACTGGCACACTAGAA -3' (#16) R: 5'-CTTACCCATAACCGAGCAGTTGAC-3' (#17)
2	F: 5'- GCAGACTGTGCTCAAACCTGGGTAG-3' (#18) R: 5'-TCTACTAATCAGAACAGAGGTTTCAG-3' (#19)
3	F: 5'-GGCACATTTACTGTGAATGAGAG-3' (#20) R: 5'-GTCTCATTTGTAGCAACTCAC-3' (#21)
4	F: 5'-CCACTTTCCACTTAGGCACAG-3' (#22) R: 5'-AGGATGGTATTTCAGCATTCTATA-3' (#23)
5	F: 5'-ATCCTAGCTAGTCTGGTAGCTGAA-3' (#24) R: 5'-TCTAGGTATCCCCGAAATCCATCA-3' (#25)
6	F: 5'-GATTCTCTGTGAGGCTTCAGATGT-3' (#26) R: 5'-GCCAACACAGCCAGCTGCACACA-3' (#27)
7	F: 5'-TCCCCACTTAACCACTTG-3' (#28) R: 5'-ACTTCATAGCTGACCATCTTCC-3' (#29)

Primers for cDNA templates:

15 F: 5'-GCCGCCCTCCTGCTGCTTCT-3' (#30)
R: 5'-GGCTTAGGATTGGGGTCTGA-3' (#31)

We analyzed raw sequence data, generated consensus sequences, and produced nucleotide/amino acid alignments (DNASIS V2.6 for Windows, Hitachi Software Engineering Co., Ltd.). Mutations were detected by creating nucleotide/amino acid alignments of reported cathepsin C sequence data versus affected PLS patients sequence data using the Higgins-Sharpe UPGMA. Numbers in parentheses are SEQ ID NOS.:.

Example I

PLS families

25 Parents of most families were consanguineous. Linkage studies localized a PLS gene in these five families to chromosome 11q14 [8]. Most affected individuals were homozygous for SSTR markers within the PLS candidate interval on chromosome 11q14, consistent with inheritance of both maternal and paternal copies of this genetic interval from a common familial ancestor ("identical by descent"). Four different haplotypes for short sequence tandem repeat (SSTR) markers spanning the

critical region were identified (Fig. 2), consistent with four independent mutations in the gene responsible for PLS.

5 **Analysis of cathepsin C**

Using RT-PCR, we found cathepsin C is normally expressed in epithelium from palms, soles, knees and keratinized oral gingiva from unaffected individuals (data not shown). The cathepsin C gene spans approximately 46 kb and consists of 7 exons. Sequence analysis of exonic, intronic and the 5' regulatory regions of the cathepsin C gene revealed PLS affected individuals from these families were homozygous for CTSC mutations that significantly altered the cathepsin C open reading frame.

Exon 6: Two affected individuals from one family were found to have an exon 6 nonsense mutation (856C->T) which introduces a stop codon at amino acid 286 (Fig. 3).

Exon 7: Three different exon 7 mutations were detected (Fig. 4). A deletion of a single nucleotide (1047delA) of codon 349 was found that introduced a frame shift and an early termination codon (TGA) 27 bases downstream. This mutation would result in a mutated protein of 358 amino acids, compared to the normal (wild type) 463 amino acids. A deletion of 2 bases of codon 343 (1028-1029delCT) resulting in the introduction of an early termination codon (TGA), and a truncated protein of 342 amino acids was identified in another family. A G->A substitution in codon 429 (1286G->A) that altered the original TRP codon (TGG) to a terminator codon (TAG) was identified in two affected individuals (#7 and #22) from two additional families. The expected truncated protein is 428 amino acids.

Although these families were not known to be related, the fact that affected individuals from these two Turkish families are homozygous for a common cathepsin C

gene mutation and also share a common haplotype for SSTR markers in the PLS candidate interval flanking the cathepsin C gene (D11S931 - D11S1311) suggests that these individuals have inherited the same cathepsin C gene mutation from a common ancestor. Additional mutations were also identified in Exons 2, 3, 4, 5, 6 and 7. Summaries of the mutations identified to date are set forth in Table I and locations are shown in Figure 5.

Papillon Lefevre syndrome is a palmoplantar keratoderma (PPK) with the characteristic clinical features of palmoplantar hyperkeratosis, and severe periodontal destruction. The PPKs are a heterogeneous group of diseases all having gross thickening of the palmoplantar skin. Clinically, the finding that distinguishes PLS from other PPKs is severe, early onset periodontal destruction. In affected individuals, the development and eruption of the primary teeth proceed normally, but the eruption of these teeth into the oral cavity is associated with gingival inflammation and subsequent rapid destruction of the periodontium. This form of destructive periodontitis is characteristically unresponsive to traditional periodontal treatment modalities, and consequently, the primary dentition is usually exfoliated prematurely. After exfoliation, the inflammation subsides, and the gingiva resumes a healthy appearance. However, with the eruption of the permanent dentition the process is usually repeated, resulting in the premature exfoliation of the permanent dentition, although the third molars are sometimes spared [5]. Destruction of the alveolar bone in PLS is usually severe, resulting in generalized atrophy of the alveolar ridges, further complicating dental therapy.

Because cathepsin C both localized to the refined PLS candidate interval on chromosome 11q14 and was normally expressed in epithelium from sites affected by PLS it was evaluated as a candidate gene for PLS.

Cathepsin C, or dipeptidyl aminopeptidase I (EC 3.4.14.1), is a lysosomal cysteine protease capable of removing dipeptides from the terminus of protein substrates, but at higher pH it also exhibits dipeptidyl transferase activity [10]. The cathepsin C gene spans approximately 46 kb and consists of 7 exons that encode a 463-amino acid polypeptide with predicted features of the papain family of cysteine proteases [11]. Unlike cathepsin B, H, L, and S, which are small monomeric enzymes, cathepsin C is a large (200 kD) oligomeric protein that consists of four identical subunits, each composed of three different polypeptide chains [12,13]. Expression of cathepsin C (CTSC) is tissue dependent [14]. CTSC is expressed in pituitary gland, spinal cord, aorta, left atrium, right atrium, left ventricle, right ventricle, inter ventricular septum, apex of heart, esophagus, stomach, duodenum, jejunum, ileum, ileocecum, appendix, ascending colon, transverse colon, descending colon, rectum, kidney, skeletal muscle, spleen, thymus, peripheral blood lymphocytes, lymph node, bone marrow, trachea, lung, placenta, bladder, uterus, testis, liver, pancreas, adrenal gland, thyroid gland, salivary gland, mammary gland, fetal heart, fetal kidney, fetal liver, fetal spleen, fetal thymus, and fetal lung.

The CTSC message is also expressed at high levels in immune cells including polymorphonuclear leukocytes and alveolar macrophages and is also expressed at high levels in osteoclasts [11,16]. The pathologic clinical findings of the PLS affected individuals studied here involve severe inflammation and destruction of the gingiva as well as hyperkeratosis of the skin from palmar, plantar and knee sites. In unaffected individuals, cathepsin C is normally expressed in epithelial tissues from sites clinically affected by PLS.

Most parents of the PLS affected individuals in this study are consanguineous. As a result, most PLS

affected individuals in each family are homozygous for cathepsin C mutations inherited from a common ancestor. Yet parents and several siblings who are heterozygous carriers for cathepsin C mutations do not appear to show either the palmoplantar hyperkeratosis or severe early onset periodontitis characteristic of PLS. It appears that the presence of one wild type cathepsin C gene is sufficient to prevent PPK and periodontal destruction in most patients. However, 1 mutation identified to date, (1047 A->G) appears to be associated with the presence of dermatological lesions. A consistent finding in the three linkage reports to date is the lack of a common haplotype among affected individuals from different families. The present report describes 20 different cathepsin C gene mutations associated with PLS. These findings suggest that the CTSC mutations responsible for PLS have arisen independently. Several mutations reported here result in the introduction of premature stop codons. While the W429X mutation encodes a protein shortened by only 35 amino acids, the introduced stop codon is 1 amino acid from the asparagine residue in the active site (Fig. 5). It is likely that such a mutation would cause a conformational alteration that may decrease or abolish activity. Additionally, we have also identified single nucleotide changes that result in missense amino acid changes in several additional PLS affected individuals from other populations, suggesting that structural alterations of cathepsin C may cause PLS.

In addition to the cardinal features of PLS, reports suggest some PLS patients have an increased susceptibility to infections [5]. This generalized increased susceptibility to infection may reflect the more deleterious effects of specific cathepsin C mutations, or may reflect the epigenetic effects of other gene loci. A variety of immunological findings have been reported in PLS affected individuals including

decreased monocyte chemotaxis, decreased neutrophil chemotaxis, impaired neutrophil phagocytosis, altered superoxide production, and decreased blastogenic response, but it has been difficult to extrapolate results of these studies. Consequently, the underlying pathogenesis of PLS has been poorly understood [17]. Immunological findings previously reported for affected individuals from the current families includes decreased PMN chemotaxis and elevated CD11b expression [18,19]. The pathologic clinical findings associated with PLS suggest that cathepsin C is functionally important in the structural growth and development of skin and in susceptibility to periodontal disease. As a lysosomal cysteine proteinase, cathepsin C is important in intracellular degradation of proteins and appears to be a central coordinator for activation of many serine proteinases in immune/inflammatory cells [11]. It is unknown if the profound periodontal disease susceptibility is a consequence of altered integrity of junctional epithelium surrounding the teeth. It is interesting that once teeth are exfoliated, and consequently the junctional epithelium is eliminated, the severe gingival inflammation resolves. A more complete understanding of the functional physiology of cathepsin C carries significant implications for understanding periodontal disease susceptibility. Identification of cathepsin C gene mutations in PLS raises the possibility of creating an animal model to study the development, treatment and prevention of hyperkeratosis and periodontitis.

Classification of the PPKs based upon histological findings, epidermolysis and localization of lesions within the skin (diffuse, linear or focal) has not been helpful in understanding the pathomechanism of disease [20]. Identification of mutations in specific genes has led to development of a revised nosology of these diseases in which PLS is grouped with the palmoplantar

ectodermal dysplasias [2]. In addition to providing insight into both normal as well abnormal epithelial growth and development, identification of mutations in cathepsin C associated with PLS will contribute to the overall nosology of the PPKs.

EXAMPLE II

CTSC Mutation in Haim-Munk Syndrome

Of the many palmoplantar keratoderma (PPK) conditions, only Papillon Lefevre syndrome (PLS) and Haim Munk syndrome (HMS) are associated with premature periodontal destruction. Although both PLS and HMS share the cardinal features of PPK and severe periodontitis, a number of additional findings are reported in HMS including arachnodactyly, acroosteolysis, atrophic changes of the nails, and a radiographic deformity of the fingers. While PLS cases have been identified throughout the world, HMS has only been described among descendants of a religious isolate originally from Cochin, India. Parental consanguinity is a characteristic of many cases of both conditions. Although autosomal recessive transmission of PLS is evident, a more "complex" autosomal recessive pattern of inheritance with phenotypic influences from a closely linked modifying locus has been hypothesized for HMS. As set forth in Example I, mutations of the cathepsin C gene have been identified as the underlying genetic defect in PLS. To determine if a cathepsin C mutation is also responsible for HMS, we sequenced the gene in affected and unaffected individuals from families with HMS. Here we report identification a mutation of cathepsin C (exon 6, 857A->G) that changes a highly conserved amino acid in the cathepsin C peptide. This mutation segregates with HMS in four nuclear families. Additionally, the existence of a shared common haplotype for genetic loci flanking the cathepsin C gene suggests that affected individuals descended from the Cochin

isolate are homozygous for a mutation inherited "identical by descent" from a common ancestor. This finding supports simple autosomal recessive inheritance for HMS in these families. As described above, we also report a mutation of the same exon 6 CTSC codon (856C->T) in a Turkish family with classic PLS. These findings provide evidence that PLS and HMS are allelic variants of cathepsin C gene mutations.

In addition to congenital palmoplantar keratosis and progressive early onset periodontal destruction, other clinical findings shared by these individuals included recurrent pyogenic skin infections, acroosteolysis, atrophic changes of the nails, arachnodactyly, and a peculiar radiographic deformity of the fingers consisting of tapered pointed phalangeal ends and a clawlike volar curve (Figures 6A and 6B). Subsequently pes planus was reported to be associated with the syndrome [24]. This was the first reported association of these clinical findings, and the condition became known as Haim Munk syndrome, or keratosis palmoplantaris with periodontopathia and onychogryposis (HMS1; MIM245010)[22]. Although the palmoplantar findings and severe periodontitis were suggestive of the Papillon-Lefevre syndrome (PLS; MIM245000)[3], the association of other clinical features, particularly nail deformities and arachnodactyly, argued that HMS was a distinct disorder. In contrast to PLS, the skin manifestations in HMS were reported to be more severe and extensive. In addition to a marked palmoplantar keratosis (Figure 6C, 6D), affected individuals had scaly erythematous and circumscribed patches on the elbows, knees, forearms, shins and dorsum of the hands. While the periodontium in HMS was reported to be less severely affected than in PLS, gingival inflammation and alveolar bone destruction are present and severe (Figure 6E, 6F). In a subsequent genetic study of this extended family, Hacham-Zadeh and

coworkers [25] concluded that the syndrome might not behave as a simple autosomal recessive trait. Based upon their estimate of the disease allele frequency in this population (0.1), the absence of the condition in other kindreds of the Cochin isolate, and an inability to document consanguinity for many of the parents of affected individuals, they hypothesized that a "complex" autosomal recessive inheritance pattern with a closely linked dominant modifier locus may be responsible for the HMS phenotype.

HMS families

Pedigrees of the reported familial relationships for the Cochin descendants are shown in Figure 7A. Descendants of the Cochin isolate studied include sibships 2, 3, 4 and 5 in the kindred pedigree originally described by Hacham-Zadeh and coworkers [25].

HMS family genotyping results

All HMS affected individuals from the Cochin kindred were found to be homozygous for all three polymorphic DNA loci (D11S1887, D11S1780 and D11S1367) flanking the cathepsin C locus. Additionally, these individuals shared a common haplotype for these polymorphic markers. These findings are consistent with inheritance of both maternal and paternal copies of this genetic interval from a common familial ancestor ("identical by descent").

Analysis of cathepsin C in HMS

The cathepsin C gene spans approximately 46 kb and consists of 7 exons. Sequence analysis of exonic, intronic and the 5' regulatory regions of the cathepsin C gene revealed that HMS affected individuals from the Cochin kindred were homozygous for a mutation in codon 286 of exon 6 (857A->G) which results in substitution of a conserved glutamine residue at position 286 by an

arginine: Q286R (Figure 8). This glutamine residue is normally completely conserved in wild type cathepsin C from at least five species (data not shown). This was the only sequence change different from the reported, 5 highly conserved, wild type CTSC sequence (GenBank Accession No.: AC011088; SEQ ID NO: 1). All available parents of HMS affected individuals were found to be heterozygous for the mutated (857A->G) allele and the wild type allele. None of the parents or siblings 10 heterozygous for the mutated (857A->G) allele and the wild type allele manifested clinically identifiable characteristics of PPK or had a history of severe, early onset periodontitis.

15 **Restriction Analysis**

The Q286R mutation creates an AvaI restriction cleavage site. We utilized this newly created restriction site to develop a rapid test to screen for the Q286R mutation. After amplification of a 465bp fragment encompassing the 3' end of exon 6 using primers: Forward 5'-GTATGCTAGAACGAGAACCGTAT-3' (SEQ ID NO: 32) and Reverse 5'-CCAATGCTAAACTTGAGACC-3' (SEQ ID NO: 33), the PCR products were purified using the Promega PCR kit according to the manufacturer's 20 instructions. Purified products were eluted in 20 µl water. Approximately 5-10 µl of purified product was digested with 5U AvaI (New England Biolabs) in a total volume of 15 µl for 1.5 hr at 37°C. Following 25 digestion, the products were separated by electrophoresis through an 1.8 % agarose gel. Amplification of the wildtype sequence results in a 465 bp product that is not cleaved by AvaI. Amplification of the mutated (857A(G) sequence results in a 465 bp product that is cleaved by AvaI to yield products of 404 30 and 61 bp. Accordingly, individuals who are homozygous for the wildtype sequence exhibit a 465 bp band. Heterozygous individuals exhibit 3 bands: 465, 404, and 35

61 bp bands. Individuals who are homozygous for the Q286R mutation exhibit bands of 404 and 61 bp. Restriction analysis confirmed the sequencing results of all examined individuals (Figure 9).

5

EXAMPLE III

Genetic Screening for PPK-Associated Mutations

10

The foregoing findings provide the basis for screening and diagnostic assays for assessing patients for the presence of mutations in the CTSC gene related to the pathological conditions described herein. A summary of the mutations in CTSC identified as associated with PPKs are set forth in Table 1.

15

Table 1. Phenotype correlations with CTSC mutations

Mutation ^a	Predicted Effect	Location	Phenotype ^b	# of Families
199-222del	Deletion of aa 67-74	Exon 2	PLS	1
445-446insATGT	Frameshift, termination at aa 157	Exon 3	PLS	1
458C>T	Threonine 153 Isoleucine	Exon 3	PLS	2
622-623insC	Frameshift, termination at aa 223	Exon 4	PLS	1
704G>A	Tryptophan 235 Stop	Exon 5	PLS	1
748C>T	Arginine 250 Stop	Exon 5	PLS	1
815G>C	Arginine 272 Proline	Exon 6	PLS	1
856C>T	Glutamine 286 Stop	Exon 6	PLS	2
857A>G	Glutamine 286 Arginine	Exon 6	HMS	1
898G>A	Glycine 300 Serine	Exon 7	PLS	1
901G>T	Glycine 301 Valine	Exon 7	PLS	1
901G>A	Glycine 301 Serine	Exon 7	PLS	1
910T>A	Tyrosine 304 Asparagine	Exon 7	PLS, RP	1
956A>G	Glutamic Acid 319 Glycine	Exon 7	PLS	1
1015C>T	Arginine 339 Cysteine	Exon 7	PLS	2
1019A>G	Tyrosine 340 Cysteine	Exon 7	PLS	1
1028-1029delCT	Introduction of premature termination codon	Exon 7	PLS	1
1040A>G	Tyrosine 347 Cysteine	Exon 7	PPP	2
1047delA	Frameshift, termination at aa 349	Exon 7	PLS	1
1286G>A	Tryptophan 429 Stop	Exon 7	PLS	3
1360A>G	Glutamic Acid 447 Glycine	Exon 7	PLS	1

1. cDNA numbering considering the initiator Met codon as nucleotide +1.

2. Phenotype symbols: PLS, Papillon-Lefevre syndrome; PPP, Prepubertal periodontitis;

HMS, Haim Munk syndrome; RP, Retinitis pigmentosa

While the mutations described in the previous examples are associated with certain pathological conditions, it is important to note that the CTSC gene contains many polymorphisms. Many of these genetic changes are not associated with the disease state. The genetic changes assessed by the methods of the present invention must be associated with the production of an aberrant CTSC protein. Accordingly, a suitable assay for diagnosing this disorder includes the step of differentiating harmless polymorphisms from those

mutations which give rise to PPKs and periodontal disorders. These include changes in the coding sequence which give rise to decreased mRNA stability as compared to wild type CTSC mRNA. Alternatively cathepsin C enzymatic activity can be compared between altered CTSC coding sequences and nucleic acids encoding the wild type enzyme. Such assays are well known in the art and need not be set forth here. See for example, McGuire et al., Archives of Biochemistry and Biophysics 295:280-8, 1992; McDonald et al., J. of Biological Chemistry 244:2693-26709, 1969; Metroione et al., Biochemistry 5:1597-1604, 1966; and Vanha-Perttula et al., Histochemie 5:170-181, 1965.

15 **References:**

1. Papillon, M.M., and Lefevre, P. Deux cas de keratodermie palmaire et plantaire symetrique familiale (maladie de Meleda) chez le frere et la soeur. Coexistence dans les deus cas d'alterations dentaires grabes. *Bull Soc Fr Dermatol Syphilis* 1924;31:82-87.

20 2. Stevens, H.P., Kelsell, D.P., Bryant, S.P., Bishop, D.T., Spurr, N.K., Weissenbach, J., Marger, D., Marger, R.S., Leigh, I.M. Linkage of an American pedigree with palmoplantar keratoderma and malignancy (palmoplantar ectodermal dysplasia type III) to 17q24. Literature survey and proposed updated classification of the keratodermas. *Arch Dermatol.* 1996;132:(6) 640-51.

25 3. Online Mendelian Inheritance in Man, OMIM (TM). Johns Hopkins University, Baltimore, MD. MIM Number: {MIM 245000}: {8/20/98}: World Wide Web URL: <http://www.ncbi.nlm.nih.gov/omim/>

30 4. Gorlin, RJ, Sedano H, and Anderson VE (1964), RJ. The syndrome of palmar-plantar hyperkeratosis and premature periodontal destruction of the teeth. *J Pediatr* 1964; 65:895-908.

35 40 5. Haneke, E. The Papillon-Lefevre syndrome: keratosis palmoplantar with periodontopathy. Report of a case and review of the cases in the literature. *Hum Genet* 1979;51:(1) 1-35.

45 6. Laass, M.W., Hennies, H.C., Preis, S., Stevens, H.P., Jung, M., Leigh, I.M., Wienker, T.F., Reis, A..

Localization of a gene for Papillon-Lefevre syndrome to chromosome 11q14-q21 by homozygosity mapping. *Hum Genet* 1997; **101**: (3) 376-82.

5 7. Fischer, J., Blanchet-Bardon, C., Prud'homme, J.F., Pavek, S., Steijlen, P.M., Dubertret, L., Weissenbach, J. Mapping of Papillon-Lefevre syndrome to the chromosome 11q14 region. *Eur J Hum Genet* 1997; **5**: (3) 156-60.

10 8. Hart, T.C., Bowden, D.W., Ghaffar, K.A., Wang, W., Cutler, C.W., Cebeci, I., Efeoglu, A., Firatli, E. Sublocalization of the Papillon-Lefevre syndrome locus on 11q14-q21. *Amer J Med Genet* 1998; **79**: (2) 134-139.

15 9. Hart TC, Bowden DW, Hart PS, Walker SJ, Callison SA, Bobby PL, Firatli E. An integrated physical and genetic map of the PLS locus interval on chromosome 11q14. *Mammalian Genome In Press*

20 10. Kirschke, H., Barrett, A.J., Rawlings, N.D. Proteinases 1: Lysosomal cysteine proteinases. In: *Protein Profile Vol.2*, P. Sheterline, ed. (London, UK: Academic Press Ltd.) 1995; 1587-1643.

25 11. Rao, N.V., Rao, G.V., Hoidal, J.R. Human dipeptidyl-peptidase I. *J Biol Chem*. 1997; **272**: 10260-10265.

30 12. Paris, A., Strukelj, B., Pungercar, J., Renko, M., Dolenc, I., Turk, V. Molecular cloning and sequence analysis of human preprocathepsin C. *FEBS Lett* 1995; **369**: 326-330.

35 13. Dolenc, I., Turk, B., Pungercic, G., Ritonja, A., Turk, V. Oligomeric structure and substrate induced inhibition of human cathepsin C. *J Biol Chem* 1995; **270**: 21626-31.

40 14. Pham, C. T. N., Armstrong, R. J., Zimonjic, D.B., Popescu, N. C., Payan, D. G., Ley, T. J. Molecular cloning, chromosomal localization, and expression of murine dipeptidyl peptidase I. *J Biol Chem* 1997; **272**: 10695-10703.

45 15. McGuire, M.J., Lipsky, P.E., Thiele, D.L. Cloning and characterization of the cDNA encoding mouse dipeptidyl peptidase I (cathepsin C). *Biochimica et Biophysica Acta* 1997; **1351**: 267-73.

50 16. Hakeda, Y., Kumegawa, M. Osteoclasts in bone metabolism. *Kaibogaku Zasshi J Anat* 1991; **66**: 215-225.

17. Hart , T.C. and Shapira, L. Papillon Lefevre syndrome. *Periodontol 2000* 1994; **6**, 88-100.
68

18. Firatli, E., Tuzun, B., Efeoglu, A. Papillon-Lefevre syndrome. Analysis of neutrophil chemotaxis. *J Periodontol* 1996;67:617-20.
19. Firatli, E., Gurel, N., Efeoglu, A., Badur, S. Clinical and immunological findings in 2 siblings with Papillon-Lefevre syndrome. *J Periodontol* 1996;67:1210-5.
20. Itin, P.H. Classification of autosomal dominant palmoplantar keratoderme: past-present-future. *Dermatol* 1992;185:163-165.
21. Haim, S., Munk J: Keratosis palmo-plantaris congenita, with periodontosis, arachnodactyly and peculiar deformity of the terminal phalanges. *Br J Dermatol* 1965;77:42-45.
22. Online Mendelian Inheritance in Man, OMIM (TM). Johns Hopkins University, Baltimore, MD. MIM Number: {MIM 245010}: {5/20/97}: World Wide Web URL: <http://www.ncbi.nlm.nih.gov/omim/>
23. Gorlin, R. J., Pindborg, J. J., Cohen, M. M., Jr. Syndromes of the Head and Neck. New York: McGraw-Hill (pub.) (2nd ed.) 1976. Pp. 373-376.
24. Puliyel JM, Sridharan Iyer KS. A syndrome of keratosis palmo-plantaris congenita, pes planus, onychogryphosis, periodontosis, arachnodactyly and a peculiar acro-osteolysis. *Brit J Derm* 1986;115:243-248.
25. Hacham-Zadeh S, Schaap T, Cohen MM. A genetic analysis of the Papillon-Lefevre syndrome in a Jewish family from Cochin. *Am J Med Genet* 1978;2:153-157.
26. Gorlin RJ, Cohen MM jr., Levin LS. Syndromes of the head and neck . 3rd edn. New York. Oxford University Press, 1990:853-854 (Oxford Monographs on Medical Genetics, No.19).
27. Hart T, Hart PS, Bowden D, Walker S, Callison S, Michalec M, Firatli E. Mutations of the cathepsin C gene are responsible for Papillon Lefevre syndrome. In Press *J Med Genet*
28. Hart TC, Stabholz A, Meyle J, Shapira L, Van Dyke TE, Cutler CW, Soskolne WA. Genetic studies of syndromes with severe periodontitis and palmoplantar hyperkeratosis. *J Periodont Res* 1997;32:81-89.
29. Weissenbach J, Gyapay G, Dib C, Vignal A, Morissette J, Millasseau P, Vaysseix G, Lathrop M. A second-generation linkage map of the human genome.

Nature 1992;359:794-801.

5 30. Lucker GP, Van de Kerkhof PC, Steijlen PM. The hereditary palmoplantar keratoses: an updated review and classification. Brit J Derm 1994;131(1):1-14.

10 31. Smith P, Rosenzweig KA. Seven cases of Papillon Lefevre Syndrome. Periodontics 1967;5:42-6.

15 32. Cohen T, Bloch N. Immigrant Jews from Cochin. In Goldschmidt E (ed) The Genetics of Migrant and Isolate Populations. Baltimore: Williams & Wilkins 1963;352.

20 33. Wilkie AO. Craniosynostosis: genes and mechanisms. Hum Molec Genet 1997;6:1647-56.

25 34. Hola-Jamriska L, Tort JF, Dalton JP, Day SR, Fan J, Asskov J and Brindley PJ. Cathepsin C from Schistosoma japonicum; cDNA encoding the preproenzyme and its phylogenetic relationships. Eur J Biochem 1998;255:527-534.

30 35. Wolters PJ, Raymond WW, Blount JL, Caughey GH. Regulated expression, processing, and secretion of dog mast cell dipeptidyl peptidase I. J Biol Chem 1998;25:15514-15520.

35 36. Mackenzie IC, Rittman G, Gao Z, Leigh I, Lane EB. Patterns of cytokeratin expression in human gingival epithelia. J Periodont Res 1991;26:468-478.

40 37. Hormia M, Sahlberg C, Thesleff I, Airenne T. The epithelium-tooth interface--a basal lamina rich in laminin-5 and lacking other known laminin isoforms. J Den Res (1998);77:1479-85.

45 38. Stabholz A, Taichman NS, Soskolne WA. Occurrence of *Actinobacillus actinomycetemcomitans* and anti leukotoxin antibodies in some members of an extended family affected by Papillon-Lefevre syndrome. J Periodontol 1995;66:653-57.

50 39. Page RC. Altman LC. Ebersole JL. Vandesteen GE. Dahlberg WH. Williams BL. Osterberg SK. Rapidly progressive periodontitis. A distinct clinical condition. Journal of Periodontology. 54(4):197-209, 1983

55 40. Page RC. Bowen T. Altman L. Vandesteen E. Ochs H. Mackenzie P. Osterberg S. Engel LD. Williams BL. Prepubertal periodontitis. I. Definition of clinical disease entity. Journal of Periodontology. 54(5):257-71, 1983

While certain of the preferred embodiments of the present invention have been described and specifically exemplified above, it is not intended that the invention be limited to such embodiments. Various modifications may be made thereto without departing from the scope and spirit of the present invention, as set forth in the following claims.

10

What is claimed is:

1. An nucleic acid molecule encoding an altered CTSC protein, said nucleic acid having at least one of the alterations set forth in Table 1.

2. A nucleic acid probe specifically hybridizable to a human altered CTSC-encoding nucleic acid and not to wild-type CTSC encoding nucleic acids, said altered CTSC encoding nucleic acid having one of the alterations set forth in Table 1.

3. The nucleic acid probe of claim 2 wherein said altered CTSC DNA has the alteration comprising a substitution of a C for a T at nucleotide position 856 in Exon 6, thereby replacing a codon encoding glutamine for a stop codon.

4. The nucleic acid probe of claim 2 wherein said altered CTSC DNA has the alteration comprising a substitution of an A for a G at nucleotide position 857 in Exon 6, thereby replacing a codon encoding glutamine for an arginine encoding codon.

5. The nucleic acid probe of claim 2 wherein said altered CTSC DNA has the alteration comprising a deletion of an A at nucleotide position 1047 in Exon 7, thereby causing a frameshift and a premature stop codon.

6. The nucleic acid probe of claim 2 wherein said altered CTSC DNA has the alteration comprising a deletion of a dinucleotide CT at nucleotide positions 1028 and 1029 in Exon 7, thereby causing a premature stop codon.

7. The nucleic acid probe of claim 2 wherein said altered CTSC DNA has the alteration comprising a

substitution of a G for a A at nucleotide position 1286 in Exon 7, thereby replacing a tryptophan codon with a premature stop codon.

5 8. The nucleic acid probe of claim 2 wherein said altered CTSC DNA has the alteration comprising a substitution of a C for a T at nucleotide position 1015 in Exon 7, thereby replacing a codon encoding arginine for a cysteine encoding codon.

10 9. The nucleic acid probe of claim 2 wherein said altered CTSC DNA has the alteration comprising a substitution of an A for a G at nucleotide position 1019 in Exon 7, thereby replacing a codon encoding tyrosine for a cysteine encoding codon.

15 10. The nucleic acid probe of claim 2 wherein said altered CTSC DNA has the alteration comprising a substitution of an A for a G at nucleotide position 1040 in Exon 7, thereby replacing a codon encoding tyrosine for a cysteine encoding codon.

20 11. A mutated CTSC protein encoded by a CTSC encoding nucleic acid, said nucleic acid containing a mutation as set forth in Table 1.

25 12. An antibody immunologically specific for the protein of claim 11.

30 13. A method for detecting a germline alteration in a CTSC gene, said alteration selected from the group consisting of the alterations set forth in Table 1 in a human, said method comprising analyzing a sequence of a CTSC gene or CTSC RNA from a human sample or analyzing a sequence of CTSC cDNA made from mRNA from said human sample.

35 14. The method of claim 13 which comprises

analyzing CTSC RNA from the subject.

15. The method of claim 14 wherein a germline alteration is detected by hybridizing a CTSC gene probe which specifically hybridizes to nucleic acids containing at least one of said alterations and not to wild-type CTSC sequences to RNA isolated from said human sample and detecting the presence of a hybridization product, wherein the presence of said product indicates the presence of said alteration in said RNA and thereby the presence of said germline alteration in said sample.

10
15
20
25
30
35

16. The method of claim 13 wherein a germline alteration is detected by obtaining a first CTSC gene fragment from a CTSC gene isolated from said human sample and a second CTSC gene fragment from a wild-type CTSC gene, said second fragment corresponding to said first fragment, forming single-stranded DNA from said first CTSC gene fragment and from said second CTSC gene fragment, electrophoresing said single-stranded DNAs on a non-denaturing polyacrylamide gel, comparing the mobility of said single-stranded DNAs on said gel to determine if said single-stranded DNA from said first CTSC gene fragment is shifted relative to said second CTSC gene fragment and sequencing said single-stranded DNA from said first CTSC gene fragment having a shift in mobility.

17. The method of claim 13 wherein a germline alteration is detected by hybridizing a CTSC probe which specifically hybridizes to nucleic acids containing at least one of said alterations and not to wild-type CTSC sequences to genomic DNA isolated from said sample and detecting the presence of a hybridization product, wherein a presence of said product indicates the presence of said germline alteration in the sample.

18. The method of claim 13 wherein a germline

alteration is detected by amplifying all or part of a CTSC gene in said sample using a set of primers specific for a wild-type CTSC gene to produce amplified CTSC nucleic acids and sequencing the amplified CTSC nucleic acids.

5

19. The method of claim 13 wherein a germline alteration is detected by amplifying all or part of a CTSC gene in said sample using a primer specific for an allele having for one of said alterations and detecting the presence of an amplified product, wherein the presence of said product indicates the presence of said allele in the sample.

10
15
20
25
30
35

20. The method of claim 13 wherein a germline alteration is detected by molecularly cloning all or part of a CTSC gene in said sample to produce a cloned nucleic acid and sequencing the cloned nucleic acid.

21. The method of claim 13 wherein a germline alteration is detected by forming a heteroduplex consisting of a first strand of nucleic acid selected from the group consisting of CTSC gene genomic DNA fragment isolated from said sample, CTSC RNA fragment isolated from said sample and CTSC cDNA fragment made from mRNA from said sample and a second strand of a nucleic acid consisting of a corresponding human wild-type CTSC gene fragment, analyzing for the presence of a mismatch in said heteroduplex, and sequencing said first strand of nucleic acid having a mismatch.

22. The method of claim 13 wherein a germline alteration is detected by amplifying CTSC gene nucleic acids in said sample, hybridizing the amplified nucleic acids to a CTSC DNA probe which specifically hybridizes to nucleic acids containing at least one of said alterations and not to wild-type CTSC sequences and detecting the presence of a hybridization product,

wherein a presence of said product indicates the presence of said germline alteration.

23. The method of claim 13 wherein a germline
5 alteration is detected by analyzing the sequence of a CTSC gene in said sample for one of the mutations set forth in Table 1.

10 24. The method of claim 13 wherein a germline alteration is detected by obtaining a first CTSC gene fragment from a CTSC gene isolated from said human sample and a second CTSC gene fragment from a CTSC allele specific for one of said alterations, said second fragment corresponding to said first fragment, forming single-stranded DNA from said first CTSC gene fragment and from said second CTSC gene fragment,
15 electrophoresing said single-stranded DNAs on a non-denaturing polyacrylamide gel and comparing the mobility of said single-stranded DNAs on said gel to determine if said single-stranded DNA from said first CTSC gene fragment is shifted relative to said second CTSC gene fragment, wherein no shift in electrophoretic mobility indicates the presence of said alteration in
20 said sample.
25

26. The method of claim 13 wherein a germline
alteration is detected by obtaining a first CTSC gene
30 fragment from (a) CTSC gene genomic DNA isolated from said sample, (b) CTSC RNA isolated from said sample or (c) CTSC cDNA made from mRNA isolated from said sample and a second CTSC gene fragment from a CTSC allele specific for one of said alterations, said second fragment corresponding to said first fragment, forming single-stranded DNA from said first CTSC gene fragment and from said second CTSC gene fragment, forming a heteroduplex consisting of single-stranded DNA from said first CTSC gene fragment and single-stranded DNA from
35

said second CTSC gene fragment and analyzing for the presence of a mismatch in said heteroduplex, wherein no mismatch indicates the presence of said alteration.

5 27. A method as claimed in claim 13, wherein said germline alteration comprises a substitution of a C for a T at nucleotide position 856 in Exon 6, thereby replacing a codon encoding glutamine for a stop codon.

10 28. A method as claimed in claim 13, wherein said germline alteration comprises a substitution of an A for a G at nucleotide position 857 in Exon 6, thereby replacing a codon encoding glutamine for an arginine encoding codon.

15 29. A method as claimed in claim 13, wherein said germline alteration comprises a deletion of an A at nucleotide position 1047 in Exon 7, thereby causing a frameshift and a premature stop codon.

20 30. A method as claimed in claim 13, wherein said germline alteration comprises a deletion of a dinucleotide CT at nucleotide positions 1028 and 1029 in Exon 7, thereby causing a premature stop codon.

25 31. A method as claimed in claim 13, wherein said germline alteration comprises a substitution of a G for a A at nucleotide position 1286 in Exon 7, thereby replacing a tryptophan codon with a premature stop codon.

30 32. A method as claimed in claim 13, wherein said germline alteration comprises a substitution of a C for a T at nucleotide position 1015 in Exon 7, thereby replacing a codon encoding arginine for a cysteine encoding codon.

5

33. A method as claimed in claim 13, wherein said germline alteration comprises a substitution of an A for a G at nucleotide position 1019 in Exon 7, thereby replacing a codon encoding tyrosine for a cysteine encoding codon.

10

34. A method as claimed in claim 13, wherein said germline alteration comprises a substitution of an A for a G at nucleotide position 1040 in Exon 7, thereby replacing a codon encoding tyrosine for a cysteine encoding codon.

15

35. A method for detecting a germline alteration in a CTSC human encoding nucleic acid, said method comprising comparing a sequence of a CTSC DNA or CTSC RNA from a human sample with an isolated wild type CTSC sequence as provided in SEQ ID NO:1.

20

36. A method as claimed in claim 35, wherein stability of said altered CTSC mRNA is compared with stability of wild type CTSC mRNA.

25

37. A method as claimed in claim 35, further comprising expressing an altered CTSC protein from said altered CTSC encoding nucleic acid and comparing cathepsin C enzymatic activity of said altered CTSC protein to enzymatic activity of wild-type cathepsin C.

30

38. A kit for detecting the presence of an altered CTSC encoding nucleic acid in a biological sample, comprising:

35

i) oligonucleotides which specifically hybridize with CTSC encoding nucleic acids having the alterations set forth in Table 1;

ii) reaction buffer; and

iii) an instruction sheet.

39. A kit as claimed in claim 38, wherein said oligonucleotide contains a tag.

40. A kit for detecting the presence an altered CTSC encoding nucleic acid in a biological sample,
5 comprising:

- i) antibodies immunologically specific for the altered CTSC proteins of the invention;
- 10 ii) a solid support with immobilized CTSC antigens as a positive control; and
- iii) an instruction sheet.

41. A kit as claimed in claim 40, wherein said antibody contains a tag.

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
1 February 2001 (01.02.2001)

PCT

(10) International Publication Number
WO 01/07663 A1

(51) International Patent Classification⁷: C12Q 1/68, C12P 19/34, C07H 21/04, C07K 14/435, 16/18, G01N 33/53

(21) International Application Number: PCT/US00/20400

(22) International Filing Date: 27 July 2000 (27.07.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/145,644 27 July 1999 (27.07.1999) US
60/165,016 12 November 1999 (12.11.1999) US

(71) Applicant (for all designated States except US): WAKE FOREST UNIVERSITY [US/US]; Medical-Center Boulevard, Winston-Salem, NC 27157 (US).

(72) Inventor; and

(75) Inventor/Applicant (for US only): HART, Thomas, C. [US/US]; 2476 Matterhorn Drive, Wexford, PA 15090-7612 (US).

(74) Agents: RIGAUT, Kathleen, D. et al.; Dann, Dorfman, Herrell and Skillman, Suite 720, 1601 Market Street, Philadelphia, PA 19103-2307 (US).

(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:

- With international search report.
- Before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments.

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 01/07663 A1

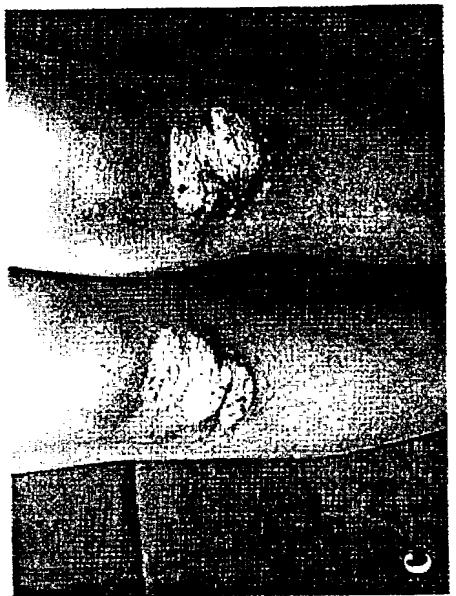
(54) Title: METHODS AND COMPOSITIONS FOR DIAGNOSING PALMOPLANTAR KERATODERMAS AND DYSPLASIAS AND OTHER PERIODONTAL DISEASES

(57) Abstract: Compositions related to mutations in the CTSC gene associated with palmoplantar keratodermas are disclosed. Also provided are methods for using the compositions of the invention in diagnostic and prognostic screening assays.

WO 01/07663

10/031889
PCT/US00/20400

1/9



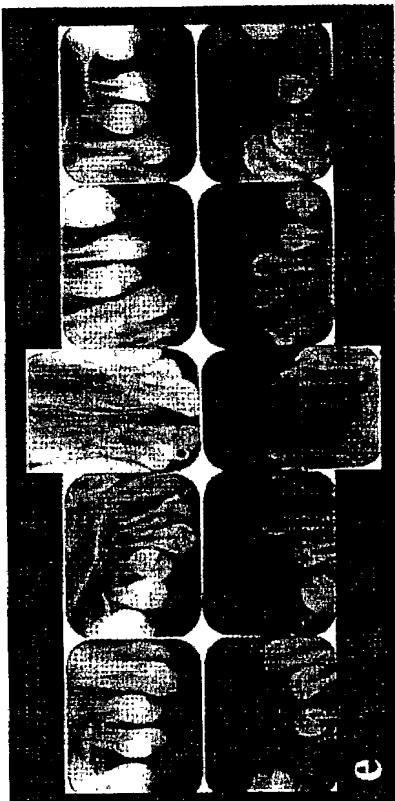
c



b



a



e



d

Figure 1

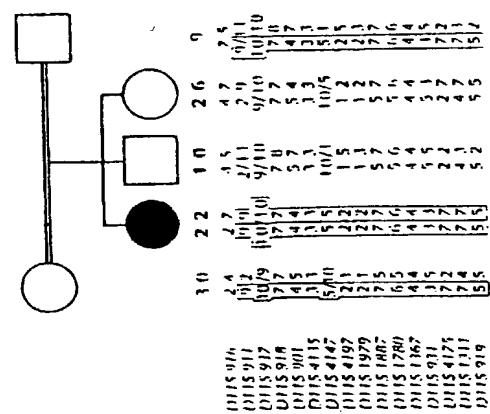
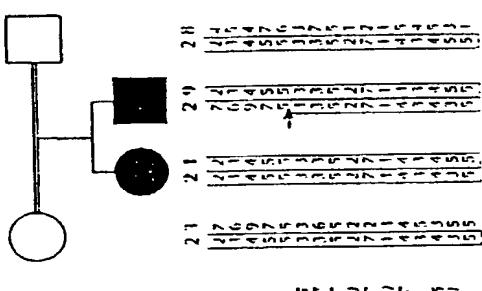
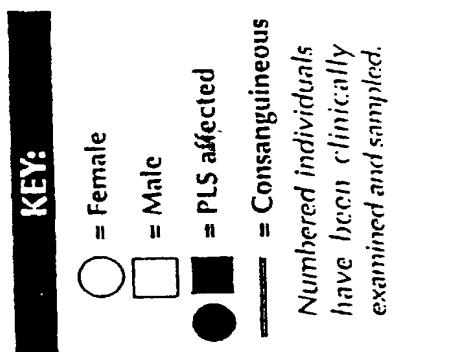
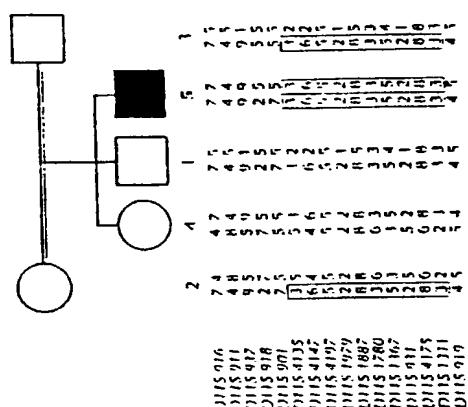
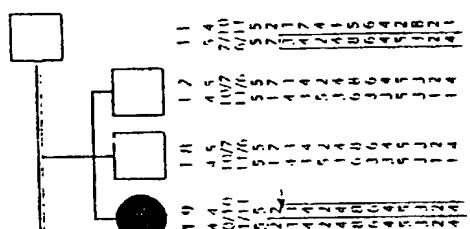
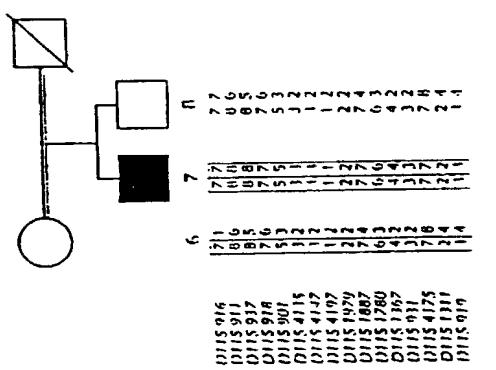


Figure 2

Wildtype sequence from nt 844 to 862 - ATCCTAAGCCCTCAGGAGG

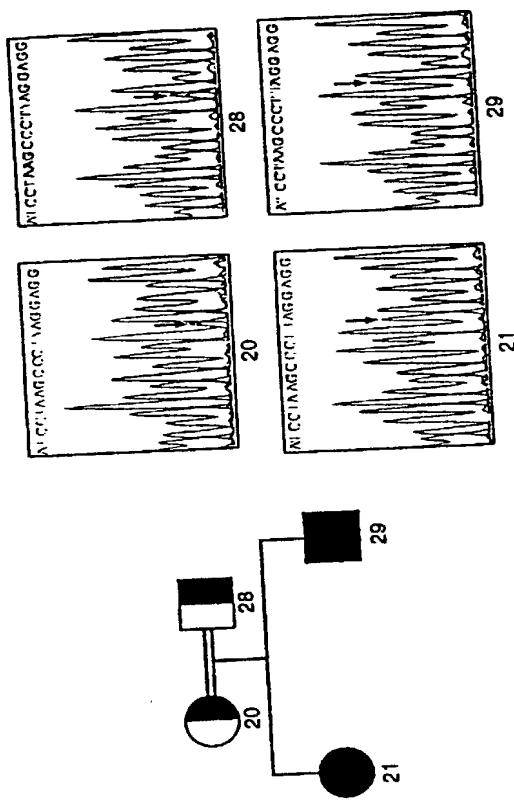
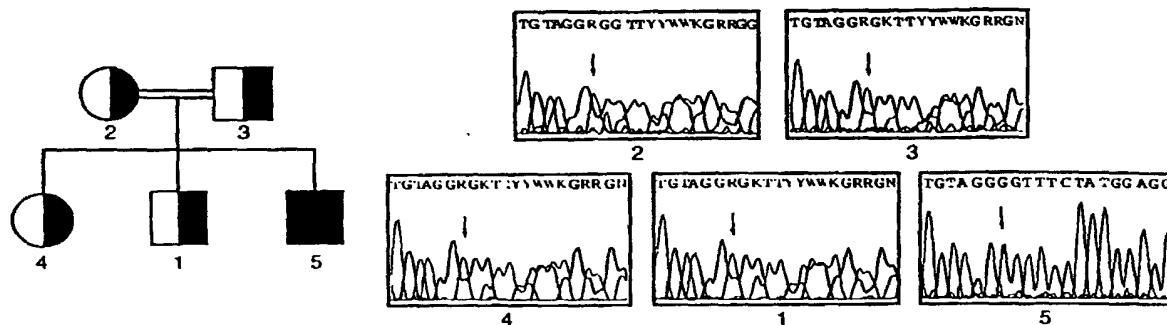


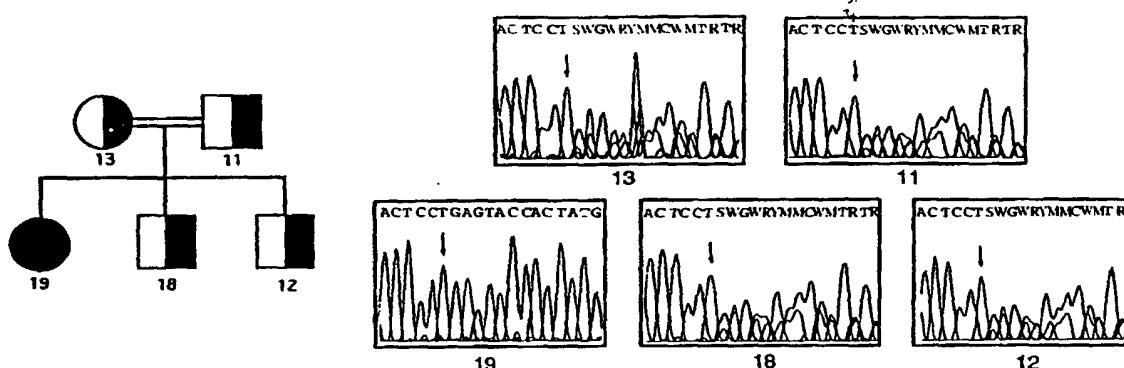
Figure 3

A

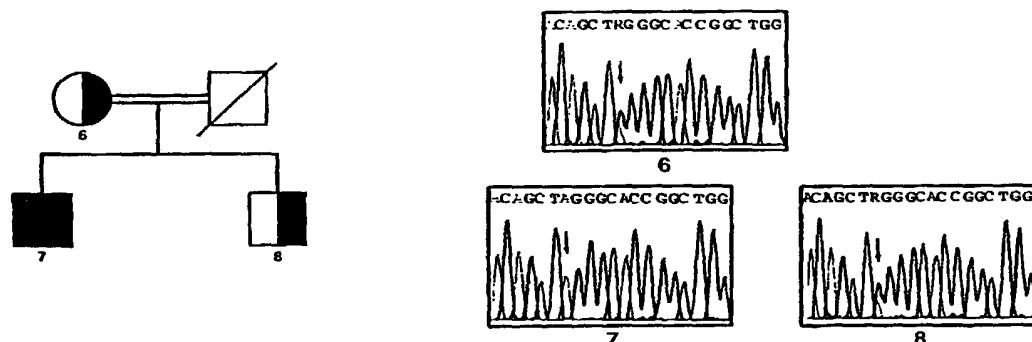
Wildtype sequence from nt 1041 to 1061 - TGTAGGAGGTTTCTATGGAGG

**B**

Wildtype sequence from nt 1022-1041 ACTCCTCTGAGTACCACTAT

**C**

Wildtype sequence from nt 1280 to 1299 - ACAGCTGGGCACCGGCTGG

**D**

Wildtype sequence from nt 1280 to 1299 - ACAGCTGGGCACCGGCTGG

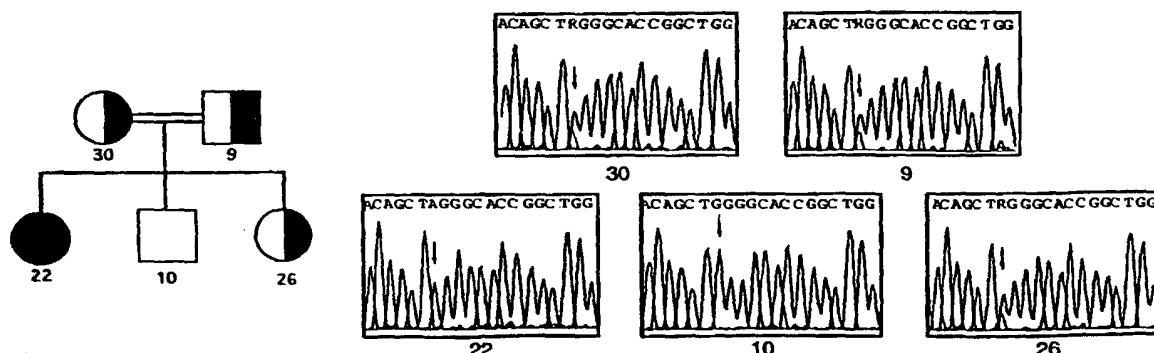
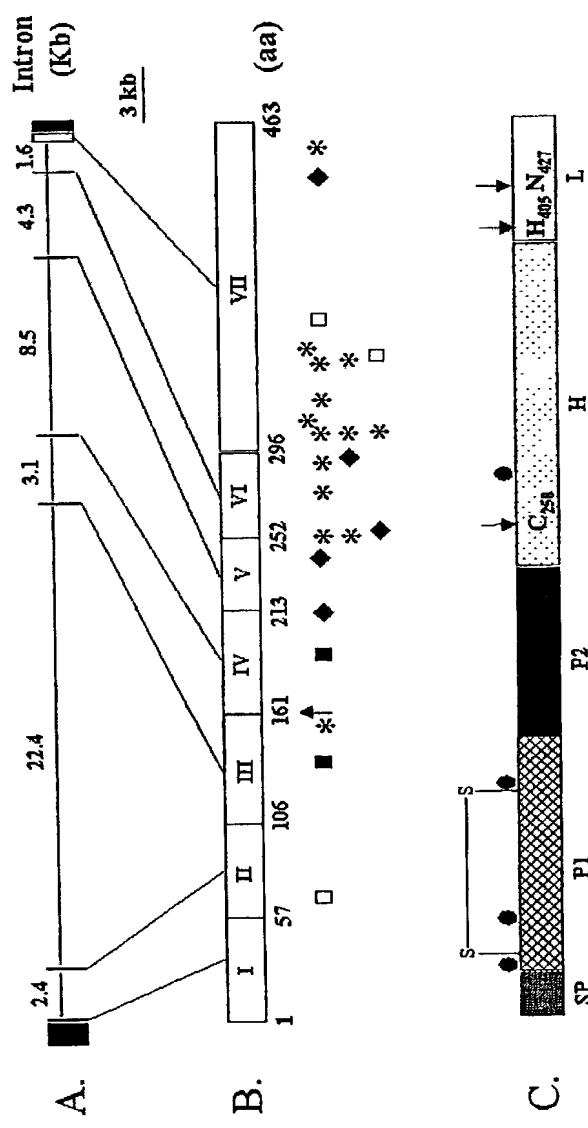
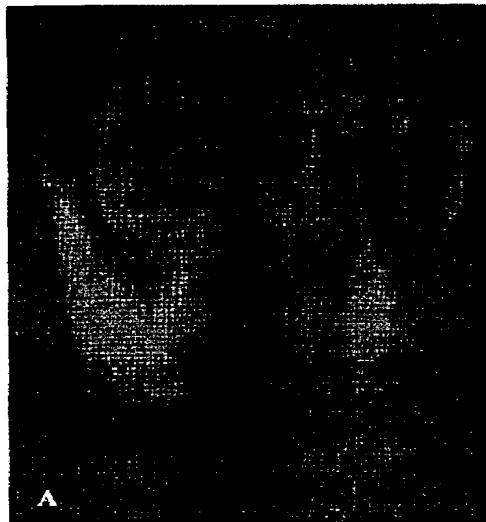


Figure 4

**Figure 5**

10/031889
PCT/US00/20400



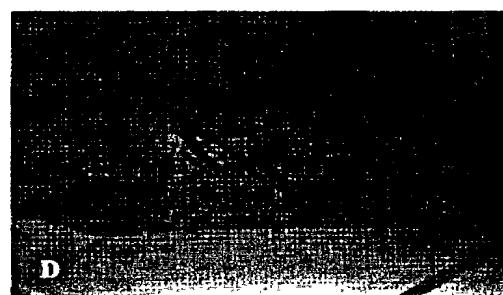
A



B



C



D



E



F

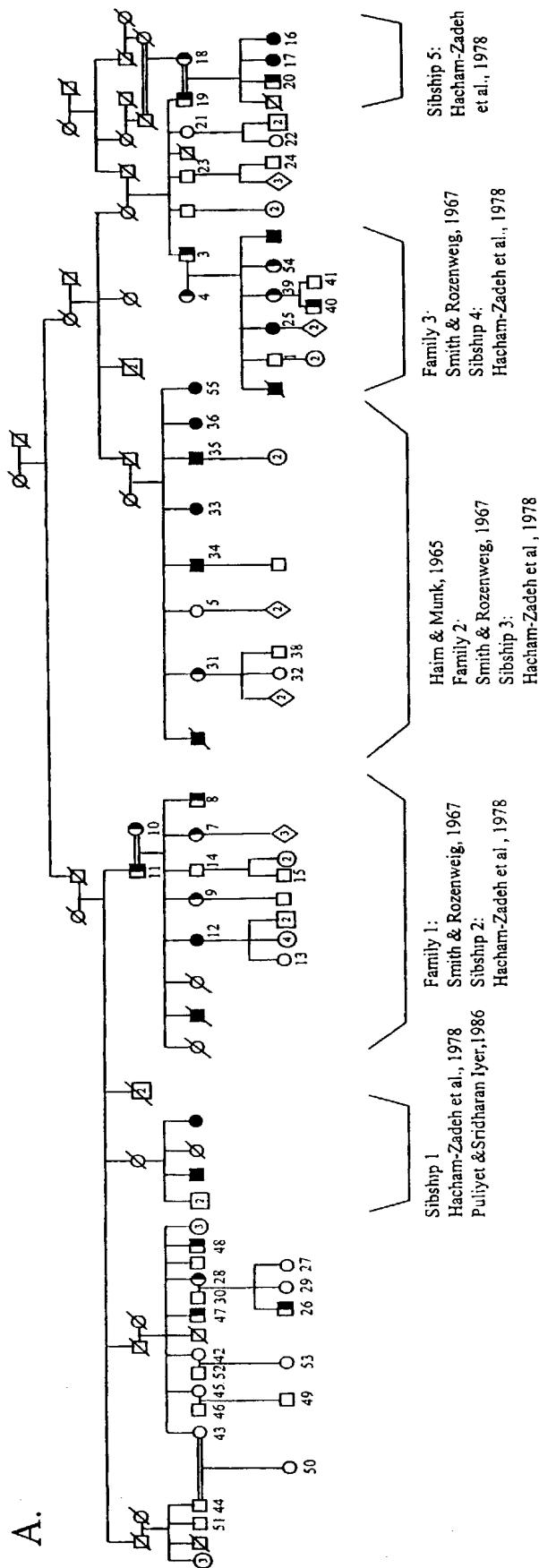
Figure 6

10/031889

PCT/US00/20400

WO 01/07663

7 / 9



B.

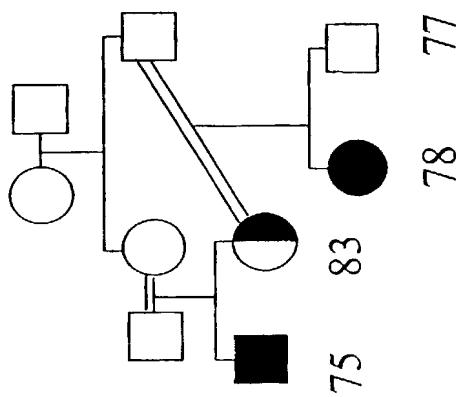


Figure 7

Wildtype sequence from nt 2114 to 2133- ATCCTAAGCCCTCAGGAGG

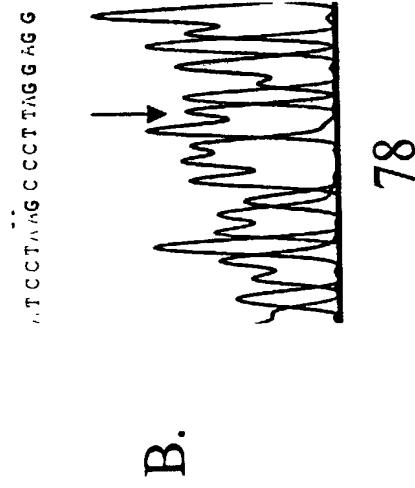
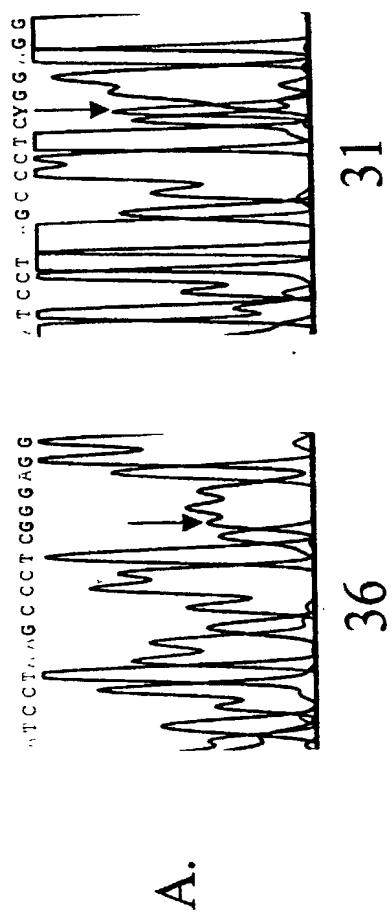


Figure 8

10/031889
PCT/US00/20400



Figure 9

UTILITY
Original U.S. or PCT D.O.

DECLARATION, POWER OF ATTORNEY AND POWER TO INSPECT

As a below-named inventor, I hereby declare:

that my residence, post office address and citizenship are as stated below next to my name;

that I verify believe I am the original, first and sole Inventor (if only one name is listed below) or an original, first and joint Inventor (if plural inventors are named below) of the invention entitled: **METHODS AND COMPOSITIONS FOR DIAGNOSING PALMOPLANTARKERATODERMAS AND DYSPLASIAS AND OTHER PERIODONTAL DISEASES**

the specification of which [check one(s) applicable]

was filed July 27, 2000 as International Application No. PCT/US00/20400 on which U.S. Patent Application No. _____, filed January 25, 2002 is based.
 and was amended by Amendment filed _____ (if applicable); (or);
 is attached to this Declaration, Power of Attorney and Power to Inspect;

that I have reviewed and understand the contents of the above-identified specification, including the claims, as unchanged by any amendment referred to above; and that I acknowledge my duty to disclose information which is material to the examination of this application in accordance with Rule 56 (a) [37 C.F.R. §1.56(a)]

CLAIM UNDER 35 USC §119(e): I hereby claim the benefit under 35 USC §119(e) of any United States provisional applications listed below:

Provisional Application No.

Filing Date
Day/Month/Year

60/144,644
60/185,018

27 July 1999
12 November 1999

POWER OF ATTORNEY: As inventor, I hereby appoint DANN, DORFMAN, HERRELL AND SKILLMAN, P.C. of Philadelphia, PA, and the following individual(s) as my attorneys or agents with full power of substitution to prosecute this application and to transact all business in the Patent and Trademark Office connected therewith: Kathleen D. Nugent, Ph.D., J.D. Reg. No. 43,047; Maria Kouraklis, Esq. Reg. No. 41,126 and Patrick J. Bagan, Esq. Reg. No. 27,643

POWER TO INSPECT: I hereby give DANN, DORFMAN, HERRELL AND SKILLMAN, P.C. of Philadelphia, PA or its duly accredited representatives power to inspect and obtain copies of the papers on file relating to this application.

SEND CORRESPONDENCE TO: CUSTOMER NUMBER 0000110.

DIRECT INQUIRIES TO: Telephone: (215) 563-4100
Facsimile: (215) 563-4044

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

SOLE OR FIRST JOINT INVENTOR

Full Name Thomas C. Hart
First Middle Last

Signature Thom C Hart

Date Jan 25, 2002

Residence Wexford Pennsylvania PA
City State or Country

Citizenship United States of America
Post Office Address:
2715 Matterhorn Drive

Wexford Pennsylvania 15690-7612
City State or Country Zip Code

SECOND JOINT INVENTOR (IF ANY)

Full Name _____
First Middle Last

Signature _____

Date _____

Residence _____
City State or Country

Citizenship _____
Post Office Address:

City State or Country Zip Code

Rec'd PCT/PTO 28 MAY 2002

HS

SEQUENCE LISTING

<110> Hart, Thomas C.

<120> Methods and Compositions for Diagnosing
Palmoplantar Keratodermas and Dysplasias and other
Periodontal Diseases

<130> WFU 99-35

<140> 10/031,889

<141> 2002-01-25

<150> PCT/US00/20400

<151> 2000-07-27

<150> 60/144,644

<151> 1999-07-27

<160> 33

<170> FastSEQ for Windows Version 3.0

<210> 1

<211> 45186

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<222> (0)...(0)

<223> n = a, c, t, or g

<400> 1

agggagatataa aagtgaataa ttggacactg ctctttttaaa atgtttataaa tctgggtggaa	60
aaaaaatgaa catatgaata ttgattttgt accagtgc aaaggggcaaa aattcatatc	120
ccaaagaaaa cggggacaca tcaggtctgt ctgttcatac actgtgtcca cagggcctga	180
cacctagtag gctcaagtggg agaaaggagc cccaattacc aacaaaagcc agggaaagaac	240
gggaggctct tacggaaaag ggtgataactt aaactgagca aggaggcacc tgaaaatagt	300
gccacctaataat aatttttggc gatcagactg acacactaga acggttcata agaccagcct	360
tctccatgtt gctagtttcc ttcttcaccc ttcttcaccc gggcaagccg ctctctctct	420
ctgggcctct tgcttttccct ctgtaacata aaaggggttt agcaatatac tctcttgagag	480
cgccatgtgt gtgcgtgcca gagggaaaac ccccacaacg ctaatatac aaaactgcag	540
gtttgcacaa aaactgaatt ctgctgtatg caaacaggca aacagcattt accagggaaac	600
aaaacaaaaat caagcacata aaaaatgtt aagagtttggaa aaacggaagg aagataagtt	660
ctcaaacacgc tgaaatagtt gatgttagct akgcgaatgtt ttccagaggaa aaaacaagaa	720
gttggttatg aggcaagtgg acctgagaaa aaagactaaa ggggaagaat agcaagtaaa	780
acagaactcc acttgtttaga tctctccctc tgcgcgttc ttccacctga cccactccct	840
tattcccccc acacccttc ttctctccctc tacgttaccc cacaggaacg aagtctgggt	900
catgtgcggc ccgcgttgc ttcttaatac ctcttttgc caccctggcc gtgcaaaatt	960
ttgaaacgtc cctcgccaaa aaaaataaaaa ataaaaaaaaaaa aaaatctgtc cctggcctct	1020
tccctagttc tgggtccagt tgcagccaa tgagggcag cgccgcgtcc caagtccccg	1080
tttcagagac ggcacacgc ctggccccc acccccaatc ccctgtgtc cagtgcaccc	1140
gccccacgggt ttccggccgc gcgttagctat ttcaaggcgc gcgcctcggt gtggactcac	1200
cgctagcccg cagcgctcg ttctcttgcatttccatc ttcttttca gctccctgtca	1260
gcatgggtgc tggccctcc ttgtctgtcg ccgcctccct gctgtttctc tccggcgcacg	1320
gcccgcgtgc ctgcacaca cctgccaact gcacatcttg tgacctgtcg ggcacctggg	1380
tcttccaggt gggctccagg ggttccaggc gcgatgtcaa ctgcctgggtt atgggttaagc	1440
cgcccgctcg gcagtcctcc gggctgtccct ttctgcctt gagccctaa cgcagcgccca	1500
cgcccaactac cgcttccccc caggcagacg ctgtgggtt gccagagcat ttgacttggaa	1560
ttcggggacc ctggggacc ttcttccccg ccaggctcg gaagttaaag ttcatctgt	1620
gagaacttct aactccacac ttcttgggtt atcttggggaa ctcaacactt tgatcaagaa	1680
cttttttatt cttcccgctt aattttgttt gctttgagag agacttggga actgcaatcg	1740

tttggttctc	cagtcgatc	tggtagcggtt	attttaaaa	tttattttta	ttttttatta	1800
ctattttact	agtgaagata	gatgagctca	gagactctcg	aggatatagc	atgaagttt	1860
ctcttttgt	tagatggtg	gaaaggact	ttctgcccag	cgatttttgt	ttgagcgggt	1920
gttgatgagt	actagaaaac	ggcttagtacc	actctgcatt	gtttcatgca	ttgcaaggag	1980
gtaaaaattt	tttaaaaaat	taataacaaa	gaaaacttaa	ctctgaacct	agtaattaga	2040
aatgcccaga	gtctgcacaa	tgtttggctc	atggaaggct	ctcaataaaat	acctagtgtt	2100
tgaacatact	ggagatattc	catatgcctt	cagataaacat	ggttaccct	agaacaaaaga	2160
acctagtgtt	gggggtgggg	tggaaaaaaag	acaatacta	gttagaagag	tcacactgtg	2220
gtctacccaa	acccttctac	agcctgtgac	ttctgcagag	tcagtaaaaa	atcagctata	2280
attttcttgt	caacagaaca	aagattaagt	catttgtac	taaagaaaata	ccttttaac	2340
tgactgtata	gcatttcata	atcctgaaca	ttgtactttt	ttttttttt	tttttgagaa	2400
tggagtcttgc	ctctgtcacc	caggctgttag	ttagtggcg	ggatctccgc	tcactgcaac	2460
ctccgcctcc	cgggtttaag	cgattctctt	gcctcagtct	cccgagtagc	taggattacg	2520
ggcacatgcc	accacgcccc	gctaataattt	tgtattttta	gtagagacgg	gttctactgt	2580
gttagccagg	atggctcca	tctctgacc	tcgtgatccg	ccgcctcgg	cctcccaaag	2640
tactgggatt	acaggcggga	gtcaccacgc	ctggcctttt	ttttttttt	ttttttaaat	2700
tttagtttga	aggttgggc	agaaagagat	caggattgc	actgcctgt	cacatgcaat	2760
ctcccatgtc	agagcattac	tctcaaacat	ggaaaaaactt	taaaatcac	aactctccag	2820
atggacacag	ctgaatcatt	ttcaggaagg	ctctgcctat	aaaattacct	ttggccttaa	2880
ttcagtaatt	aaaggcaccc	acaggtctga	gcctctatct	atgaaggtt	aacgtatcat	2940
tcctatcagc	actaatttgg	tttataaggat	acagacctt	agttgtgt	ataaaactttt	3000
gaaggattca	gattatgagg	ttctaaactg	taggccttt	gaggagaaag	gtaccccatt	3060
tcttctccga	atagatcatt	ttgtgtcttc	tcctgggtt	agcattgt	cttctttagg	3120
atctgatagt	ctgttttattt	atcttttgg	tgtcataacct	tttattcttgc	catttcccac	3180
ttttgacttag	catttgcct	tttctccgtt	tctggaagcc	tgtatatttc	acatttcccc	3240
atttctcctt	ttgttcagtgc	aaaaaaattt	ttaagtgttg	ggatgtctg	gaaaggtaact	3300
gtagtttgg	gggttccct	gcctgctggg	ctgcatacg	tatctcttt	ttttgagaca	3360
aggctctact	gtgtgcccc	ggttggaaatg	cagttgtgc	atcatagctc	actgcagcc	3420
ccaaactctg	ggctcaagca	atcctccac	ctcagcttca	gccgctcaaa	tagctggac	3480
tacaggcaat	gcaccacca	aaccagcgaa	ttttaaaaaa	ttattttgg	agagacaggg	3540
tctcaactatg	ttggccaaac	tagtcttgc	tttctgacct	cagccccgca	aagtgtctgg	3600
attcacaggcg	tgagccactg	cttctgttt	cattatctca	aattttttaga	gcctaaacttc	3660
acagtttgc	ctcgaggcag	gaactgttt	aggcaacaag	atggagcctc	tgtttttttc	3720
actaggcaga	ctgtgtctaa	actgggttgc	atgaaaggaa	agcagcacaa	ttaaaatttg	3780
aattggggga	ttcttcccc	tatttggaaatc	caaataattt	tcctttatttgc	tttttttttt	3840
ttttttctt	aggaccacaa	aaaaaaaatgg	tagtgggttgc	ccttcagaag	ctggatacag	3900
catatgatga	ccttggcaat	tctggccatt	tcaccatcat	ttacaaccaa	ggctttgaga	3960
ttgtgttgc	tgactacaag	tggtttgcct	tttttaaggat	tagttttgtt	ggaagtggaa	4020
tttacattt	caatgatttgc	atatctgaaa	cctcttctga	tttagtagacc	ctcagaattt	4080
taattttaga	ttaagaagat	gaccggaaatt	gacaccactc	ttcccaagag	tagtagtagt	4140
tggtataatt	ttgccacttt	atttaaaattt	atgatttttag	tagtttaca	aaataccagt	4200
gctacatttgc	aatatgtata	aattatgttt	aaaatttaca	ttttggtaag	tatgactaaa	4260
ttcttaatttgc	attttccttgc	ttactaccac	tttattttctt	aatgttgc	ca tagtcat	4320
gtttgttctt	aaatctgttag	gaaagataga	gagattacac	attttttttgc	tttgcagtt	4380
ctatgtgtc	cttcttctca	ctactgttt	gctgagggtt	tgataggcct	aaatgattca	4440
ttatctttaa	tgtactaaat	atgttgcata	attttttttgc	ctaaactaac	agaaaagagag	4500
aaccttaggg	ttactcccttgc	aggctgggtt	aagtggaaagg	tagccaagtc	aacccagctt	4560
gtttcttcttgc	ctcatttagga	aagaactatttgc	tttcatttc	ataacacact	ttttccaatt	4620
gcaaacatata	tcagggtttaa	aatagtttag	cacaaatttgc	agcccatttc	atttgcattt	4680
cacaagctgg	aacttttcttgc	gtaagctaaa	tattaaatgg	ttcaagttaa	ttggatacatt	4740
aaggctgaaa	ctaggcggtt	ctcattatac	ataagatata	aattaagaca	gacttttca	4800
tggtgaaagg	tttacagcc	ttaaaaacatc	tgggaagaag	tgggaaagta	ggaataact	4860
ctgtttaata	tgataaaaaga	caaagccatc	acaaaggctt	agttctaaac	ttgttataat	4920
ttctctatgg	agttgtggtt	tgtcacaagg	ttatggcggt	ccaagcaat	ttacaatatt	4980
tttttagaata	ataactccca	gaaatattttt	aaaataagg	accttttctt	aatatggaaa	5040
aaaaaaaaatgg	gaaatagaga	ggaagaggtt	gcctttcc	aactggattt	gtgtgtgt	5100
tatgcagctg	gtggaaacca	gtttgtccat	gctagcttttgc	gatggccacc	agtccactgt	5160
atttagtggaa	accatgtttgc	aaccatgttgc	taactatttttgc	tagtacat	tccagatgt	5220
acgcatttttgc	tttggaaaaaa	acatacttac	aatagatatct	atataaatttgc	ttaaaaattgt	5280
aatgtatata	ttgtatatttgc	aaccgttttttgc	taatcttttgc	tgtatacg	tttgcagcc	5340
ggcctttgc	gacccttttgc	agttacttgc	tcacagccct	tcaaagg	tttgcattca	5400
gattaagaat	caattgcatt	tggaaatgcatt	gtgaaagggaa	agagatatttgc	aaggcaatttgc	5460
ttaccatgttgc	ccataaaccttgc	tgtacataat	ttttgttgcatt	tttcttcc	tctgtcttca	5520
tcccttcttgc	tttgacacac	aatagacacc	tagtcaatttgc	attacaaaaaa	aatgaatga	5580
atgaagtgg	attcagttgg	gaaataggttgc	aaataaaatttgc	tttaggagat	gaggaatagg	5640

taaaaagaga	caagtacata	gtttattctt	ttgacttaga	aaactttga	ttcttaaatt	5700
ctgcagaatt	ggagaaaactg	gtggggaaac	ttctaaaatc	attatttaat	taccagagat	5760
gtaatagata	tagacaaaag	cagttttctt	cttttattat	tttttcatca	gttagttctt	5820
agcttaataa	gtagtc当地	gctggtaggg	acagagggaa	ttagctggtg	gctgaatgag	5880
gaatttgtatc	acttttgtg	aatcacgggt	taagcacatt	ttgtgttttgc	ccattgccta	5940
agaacattag	tcacattagg	tcaatagaaa	atcactttt	aaagccaaat	aaagttatata	6000
gtgttcccaa	ccatgagttg	gaaagaatta	atatatgtc	tgttgagggg	tagaaccctg	6060
cctaattcata	tggttcttgg	tggcatttgc	cgaatcotta	tttttcattt	agaataaaca	6120
atagaaaaaa	tactctggcc	ctactgattt	cagatatgt	ctattttaaa	gtgcccattt	6180
gacaaaacca	ttatcaggggc	catgtttctt	tttctgcag	aaaaatcaac	cactctggc	6240
agttagttagg	tcttatgaca	agcaccataa	tttctttagg	cagagtagaa	tataatagga	6300
tacttctttt	tgaaaacttaa	tataatcagg	tagttccaga	taaacatagc	ttgaaagtg	6360
ataaaatacc	atgttatttt	agtaaatcca	attgcaagag	tgatggggaa	cagagtttaa	6420
aaacttaaga	aagatattaa	gatggagttg	actttgaata	ataaaagtcat	ccactgttga	6480
tgggtgacat	ttattaatac	agaaagttt	cagattttac	cataagcatc	agggtatttc	6540
ctgcagctgg	ggyaaaccgtg	cttggaaagga	tgcgtactc	aaggaaaaca	caagccatg	6600
taataaagtat	tgcatgtgag	aatttttctt	atagaatttag	aaagcatctt	ttacattaa	6660
atttattttt	gtaaaaagag	aaaacatcaa	aacctggata	gtattttctt	ttcaaagagt	6720
cttacataaa	cgaaaacata	ctcaaccatc	tgcattaca	gaatatttttgc	acaattttctt	6780
gccatgtta	cctggccatcg	tttgggttat	cctacaaatc	aattggattt	tcacgcctt	6840
caactgactg	gaaccctaca	acttgc当地	ttttatctt	tttatatatg	tttcagatata	6900
gcttgaagta	gatttttctt	tatttttctt	getgtcgag	tttgggttgc	agttggata	6960
tgcagagttat	ttgtatgtt	tgacatata	agggttttgc	catggttctt	ggctcataac	7020
tcactcccaa	gacccttgg	acagaaacca	gaatctctt	ctctgtatctt	ctccctaccct	7080
ccttcatct	gcccaactgca	gaactctaa	ctgattatgg	tttcttaagac	cctcataccca	7140
gagagtattc	tgccccatata	catagcagaa	ggaacactgc	acagagacac	caagaagaat	7200
ctgaacacagac	aggccttgc	aggttttagat	catgtcctt	taacctaatt	atattttaa	7260
atgggttatcc	atgtttaat	catgtgtatt	caatgaacgc	tccataaaag	cccaagaaga	7320
acagatttga	gggaattctg	aaggcgctaa	catgttagggt	ctgacagggaa	ggtgaagaag	7380
aactcatcat	gcttggaggg	ttgtccacccc	taactccgca	gggacagaag	ctccctgtgc	7440
tggggaccctt	ccagaacttg	ctctatgtat	ctttcaagt	gcctgttat	ttatattcctt	7500
tcaaataatcc	tttgtatataa	atcataaaca	tgtttccctg	agttttttagt	ccactctgtc	7560
aaattaatgt	aacccaaaga	gggggttatg	ggaaccccaa	tttgaagcca	gtccatcaga	7620
agttctggag	gccttggactt	gagactgtg	tctaaagtgg	gcaggcagtc	ttggggacta	7680
agccctcaac	ctacgggatc	tgaaactgtc	tcctgttgg	tagcattgg	gttgtactgg	7740
agggcactga	gcttgggtct	gctcagaat	tgattgtctt	cttgctgg	gggagaactt	7800
cctacatatt	ttggggtca	cgaagtctt	tgtattgtt	gttgggtctg	agcttctatt	7860
gcttagaaact	tgcttattta	agttctgtt	attcagggtc	tatctgaaag	agaaacattt	7920
tttatgtatt	gagatttcta	agccatttta	aaacttggtt	ttaacagctt	gagaattttgg	7980
gggtatagta	gggatggaga	tacatata	tggtatgtac	ttcaagctag	ataaaagttt	8040
ggaaggaata	aaagtttgc	atctaaagtt	tttgcata	tttgcata	caggaggta	8100
taggtatgtt	ttttaaaatat	tttttcttgc	caggcacgg	gactcatacc	tgtatccca	8160
gcactttggg	aggccgagac	aggcgatcac	ctgaggcag	gcgttcgaga	ccagtcgtgc	8220
caacatgaag	aaaccccg	tctactaaaa	ataaaaaat	tagtgggtg	cggtgacaca	8280
tgcctgtat	cccaactt	caggaggct	aggcaggaga	attgttgc	cccaggaggc	8340
agaggttgca	gttgcggag	atcacaccat	tgcactcc	cctgggtgtc	tcaaaaaata	8400
aataataata	aaaaatgtaa	agattttttgc	ttactcagca	tcctccaggc	atttttattat	8460
ctgagactt	tatgggagtt	gcatattaca	tttggggccc	actcagggtg	gtgggtatct	8520
aagcatttga	aataaacctt	tgttaactaa	taaggagtaa	tcaggcctgt	ggcaagatgg	8580
aaacagtctt	agaggcattc	aaatccaaat	ttcctttaaa	acactgggt	gccaaaaaca	8640
aaagacaata	ctatctacag	gccagttct	aagattatca	gattttgt	gcatttacca	8700
tttcatttgc	tttggc	tttagccaaat	ttgcacttct	taaaagtacc	tgcaggcaat	8760
ctcctatata	aaaacacaat	gcaggctagc	tttggctctg	cctttaattc	cagcacttgc	8820
agaagtttgg	gacttagcgt	gccaacgtgg	tggaaacctca	tctccactaa	aaataaaaaaa	8880
attagccagg	catagcagcg	cacgcctgt	gttcccaggta	tttggggaggc	cgaggcagta	8940
gaatcactt	aacgctggag	acagagcatg	gagttagtgc	agattgcacc	actgcactcc	9000
atcctgggt	acagcgttag	actctgtctc	caaacaaaac	aaaacacaca	cacacacaaa	9060
gtaacaacac	gaaacagaat	actgtgaaaa	tgtttaattt	tgtctgactt	tatcatgtgg	9120
ctggatatgt	gattttttt	tcttcttatt	gctttatgt	actttgttca	tttttaatga	9180
tgatcatgtt	taaagcttct	ctgtgttagc	ttctctccac	caaattgc	agagacagga	9240
agtctgtt	aaacaaactt	gctccaaaaaa	atgacactt	gttgaatagg	tttttttttt	9300
tttttttttt	tttttttttt	aatgggtc	tagtctgtc	tccaggccc	tcgctcctt	9360
caccccttgg	gtttaagaga	tttgccttcc	tcagcctcca	gagtagattt	gattacaggt	9420
ccccggccatc	acgcccagct	aatttttgc	tttttagtag	agatgggtt	tcaccatgtt	9480
gaccagactg	gtttgtact	cctgacccca	agtgatccgc	ccgcctggac	ctccccaagt	9540

gctgggattt	cagggtgttag	ccaccacgcc	cagcctgaat	aggcttctta	acctactgtt	9600
tctcatttt	ctttctctga	ggcagtaaaa	gaaactgact	ctaaaaggga	gcagtagaga	9660
aggaactca	attttatttt	gaagattaag	ctactcaagg	gctgagaaaa	tatgttagagg	9720
ggaagttagac	tcattatggc	tggaaatttt	atttggtat	agtgaagcaa	atcttagggc	9780
tttctaattt	agctcttggt	tgaaaggctc	caatcttaat	agaactaa	gctaaaaaaaaa	9840
atgaccatca	gtgttctaa	agcaagtgc	tactaaaaac	aagaacactt	tgggaggctg	9900
gggctgggtt	atcacctgag	gtcagaagat	ggagtccagc	ctcaacatgg	tggaaacccct	9960
ctatactaaa	aataaaaaaa	gtagccccgt	gtgggtgtc	acgcctatag	tctcagccac	10020
ctgggaggct	ggggcagaag	aatcgctga	acccgggagg	tggaggttgc	agtgagctga	10080
catttgcca	ctgcatttca	gcctgggtg	cagagtgaga	ctctgtctca	aaaacaaaaac	10140
aaacagaaac	aagaactcta	tgttgagaaa	tccatactag	agggttaat	ttctcacttt	10200
gtgtgctagt	gattaataaa	tttcaagctt	atcacacagc	cagaaatgtc	gcctgtgttt	10260
ctacaataaa	agattggga	atattgtgac	atttttcac	tgagccttct	gggttcattta	10320
ttaagatata	acaattttag	aagacttatt	gagataggtt	tacttttaa	aattttggatt	10380
caatatata	agctcagcat	tttgtcttt	ttttgtttt	ttttaaacaa	cttgggttt	10440
acttgtaa	gtatttcagt	ggaggaaaaa	tatcagaagt	ttagcactga	gtgeectgcct	10500
aactatttt	tgtcttctt	catgtatgt	cttaatctca	acataatgac	tcaagtttacc	10560
attgtaaact	ctttgttagg	tgtcactgt	tttatcgtc	atttggtcat	gcagctggga	10620
actgtgggg	tatgtattt	gccacatctg	aggaacaaaac	tgggtgagct	ccattgaaaaa	10680
atttctaact	ttcttattaa	aagcaagctt	cctcaactcc	tgatggctta	tgcaaattaa	10740
tgccttattt	ttcatcagag	gtgtactgga	gagattatgg	ttttcttatg	aatttcttgc	10800
ttgagcctgc	ttgtctgtg	tttcttaagg	gtctgcacta	ttttatgtga	gtaattttct	10860
acttttattt	tttctgtttt	cacatgtcaa	atcagctctc	ccaagaatgc	tacttgaac	10920
ctaagtagac	gtgaaccgaa	aagggttaaag	acccagctaa	aaaaaaagggtt	gactcaagtt	10980
cagttcacat	tcgtaaaagta	gttaggcgtc	cttaggttgc	ttattttttt	tttggaaaaga	11040
atccccta	agagaaaacat	gtgaggccac	agcagcttag	atctgttttc	cacagagaag	11100
gtggctttt	cagagaaatt	gacatactca	tcacttatct	gacatgaccc	agctttgtaa	11160
aactggctc	tattaaaata	gtcttaatag	actattcact	gaggaggaga	aatcttatt	11220
cactcttac	attctcttca	cattttcaga	atagatgttt	aaatcattgc	tcacactgga	11280
tccataaaatg	tctaaaatgt	tgatgaagaa	ataggtgatt	agagagaaaa	attaaatagag	11340
acttaccctt	cttgcattt	taacataata	ttctttcccc	ttttccttcc	tctgttactt	11400
ggctcttaaa	taccagaagt	gagatgatgaa	aaaggaaact	gggaacaagt	atgaaaagca	11460
ccataggtt	atcttatatt	agcattttcc	aaactttata	atgaaccagc	agtgaccac	11520
acatctgc	ggtaggaatt	gctcatcgt	tgtgtcgct	tatctgtta	agcttcagca	11580
atgttccat	cagccactaa	taacagatca	aagtgagcat	tagggttga	attagtaagc	11640
tgtttcttct	cagttttttc	ttgttagttg	ataaaatata	aatgtattaa	atagttttct	11700
ttatttcaga	actttcaga	gtctgtata	tattgtatgg	ttgttagctt	ggaggaagat	11760
gcaataggaa	actttccca	gataggttca	ctatTTTTT	tttccacgaa	aaataagctg	11820
ttctcaaaat	acagtttaca	aaatttttac	cttaactctt	cacttttct	cctagttagg	11880
gagaccgctc	caccagttaga	aaagataaaac	cctggtaatt	tgttgtgtt	atgggataaaa	11940
tagcttagta	cctagtcatg	tggattcagg	cagcactgag	cctaaattaa	agtttgcag	12000
gtatacatgt	taatgttatct	aagttactat	atttagcctg	tttcttaagt	atgtttcaga	12060
aacatattcg	tttttttcag	ttgcagttac	cttcagatgc	atgtgtttct	aaagcatgtt	12120
ggttgcattt	gcagtcattt	ttgtcttaagc	atttagctt	agaatggcat	ctttcctgtg	12180
aatgtcttaa	cattacaaa	aatataccag	gatcttaat	atcagtgctg	gtatTTTTT	12240
ttttttttac	ttaaagaaaac	tgtatgatt	aaatattaag	agacaatatg	atccttgcgt	12300
gtttgttacc	ctagtttttt	ttgttcttgc	ttttttttat	agagcatctg	ttgaggggact	12360
cttttaaaac	cacacccatg	aacagacgtt	ggggctaaga	gacagagcag	cctgcacac	12420
tgtggaccc	cctgttagcag	ctagcaaaagg	cctctagcag	ctacagtccc	ttctggagtc	12480
tttatttgc	tgcaaaaatgc	aaaggagttcc	ttgtgaccta	cctccaaaggc	agctgcccctc	12540
ctgaacactc	ccttggaaaa	cagtaaacat	cattttggaa	tgtgaacaac	cagagactac	12600
acaggagaaa	ggaaaaaaa	attctgaaga	tgcaaaatct	tgggtggctt	caccgttca	12660
tttttaata	aaaggaacaa	tatacaacac	gttggctttt	ttctctttt	aaatcccttc	12720
tattacagt	atttttttct	aagattgtca	ggatttgaag	tgtatgtttt	tttttattca	12780
cagcgtaaat	tttattcaca	gtttaactgt	ctgcctgagt	gtctttctt	tctctaatta	12840
ccttgaggaa	cccaagagcc	tgttgttaagg	agaaaaataag	gcccttggat	ctcttgagat	12900
tcacagat	aagttattga	agggaaagatg	gtcctatgga	ggacatattt	aaagaaggg	12960
aaaaagaggg	tttctcagat	atgtcagact	gctatagtat	acttctacag	atttagacc	13020
tccagtagct	ctggccagaa	agatggtac	gtaaaacaccc	tatTTTTT	tttttctttt	13080
ttcatttagt	acaagcttt	tgcttaagaa	ttgacataact	ataagctaca	aaagttctgt	13140
aaagtagata	taacttagtt	cattttatag	atagagaaaa	ttaatcttt	acagtgcata	13200
gctcacagag	tttcttaact	taaaaatgcta	gaacttgc	ttcaagccta	aagacttct	13260
tggggctaaa	tagtaaaaaa	agccatttca	caaataagta	aatggtattt	agaggcatat	13320
ttggatttcc	tggtaaattt	cagtcgtgt	gcatcatgaa	tattagttt	atgttgcatt	13380
ggctcatgtt	gaagttttaa	gagaagaact	gccttgaagc	ttagggttcc	ttagcttattta	13440

ggctactgac	tttcttgcc	aaaccagggt	ttttcattg	aagaccaaaa	cttacccct	13500
ccttcagtt	gtagtttgg	aatttgtaga	agagcttgc	aaacttcaa	ttaagtacaa	13560
actaagtgtc	atagtcaa	at ttaact	taattacagt	attgttcaac	tgattgttat	13620
cttctagctc	tttcctgccc	aataatggc	ttgtttctg	ctctgttg	tttagagctga	13680
cttcttcag	ctttggtaag	cctgaaat	tgggttat	ttaattcat	attgtctggg	13740
tggactttcc	tctcttgat	ttctgctga	atagaagaat	tttctctag	agagtagttt	13800
gtcataccta	ctctgttat	tcagatgact	cttggatga	tctgagaggt	atactgttct	13860
gttatttgc	gaagaagat	ttcagaaga	tgaattaaga	gtacagtgg	ctgctccac	13920
ctggaaactt	ttatctatct	cacctctg	cctgataaaat	tctttatcac	tcaggacctt	13980
gatgacgctg	ctctctgaaa	ccctcccg	ctctcttat	taccgtgaga	aacatcagaa	14040
cttgggtcc	cattgcatat	cgcaggta	tctgtttca	tgccatgt	taatggagtg	14100
attgggttagc	atgtttcat	ctcttccag	attgaaaatc	tgtatttctc	cctgtatatc	14160
ttcaaacacct	aatgcacata	gaactttgt	ggtacctgg	aaatgcacca	cagtttctt	14220
ttcttttgc	agactttca	caagtattac	caacttacaa	agaattaatt	ttttaggatt	14280
ctagaaagac	aaatcaggaa	tggtccata	tacatcttt	ttgattccct	gctctaaaga	14340
atattatcg	gttacctcc	tgcagagtt	taaaagaatt	gcatattca	agctgactt	14400
caggatgtaa	atataaccaa	agcaactgat	atgaaaaaaa	tatattcaat	ggcattccct	14460
gattttctc	taggtgttt	tattgtttt	gggtttat	ttaagtcttt	aatccatctt	14520
gagttatttc	ttgtataggt	ataagaaagg	ggtccagttt	taatttctg	cgtatggct	14580
gccagttctc	ccagcaccat	ttttaataa	ggaaatcctt	tccctattgt	ttttaggtt	14640
acggtttgc	aaagattaga	tgggtgtt	tgtgtggct	tatttctgag	atcttcatc	14700
tcttccactg	gtctatgtt	ctgtttttt	accatgctt	tttggttact	gtagccttgt	14760
agatagtagt	gaaagatagc	atgatgcctc	caggtttgtt	cttttgctt	aggattgtct	14820
tggctatacg	agttttttt	tggttctata	tgaattttaa	aatagttct	tctaattgt	14880
tgaagaatgt	taatggtagt	ttaatggaa	tagcattgaa	tctgtgaatt	gcttggcca	14940
gtatggccat	tttcatgata	ttgattttc	ctatccatg	gcatgttaac	ttttccctt	15000
cgtttgtgtc	ctctctcatt	tccttgag	gtgggttgg	gttctcctt	aagagatct	15060
tcacttcttc	tgtattcc	gatattttat	tctctctgt	gctattgg	atgggagttc	15120
attcatgatt	ttgcctctg	cttgccttt	gttgggttat	aggatcc	gtgacttct	15180
cacattgatt	ttgtatctg	agacttacc	gaagttgtt	atcagctt	gaagctttt	15240
gctgagatg	atgggtttt	ctagatata	gatcatgtt	tcttcaa	aagacaattt	15300
gacttcctc	cttccttatt	gagta	gttccgtt	tcttgcct	ttgcctgc	15360
cagaactccc	aatactat	tgaataagaa	tggtagag	ggcattc	gttctgtgc	15420
agtttca	gggatgtt	ccagtttgc	ccatccat	atgatattat	ctgtgggtt	15480
ctcataaaaa	gcttattt	ttttagatac	tttgcctt	tacctat	attttagattt	15540
ttaacatg	agcgatgtt	aattgtatcg	aaggcctt	ctgtgtct	ttagataatc	15600
atgtggttt	tgtcttagt	tctgtttat	tgtatgt	cgtttattg	tttgcata	15660
ttgaaccggc	cttgcatct	ggggatgaa	ccaactgac	tgtgttagat	aagcttttgg	15720
atgtgtct	ggatttgg	tatcgtatt	tcattgagat	ttttgcgt	gaagttcatc	15780
aggatattt	gactgaagtt	ttctttttt	tgtcgat	ctgcccagg	ttgggtatc	15840
gatgatgtc	gcctcataaa	atgagttag	gaggatccc	tcctttca	ttgtttggaa	15900
tagttcaga	agaaaaggta	tcagctctc	tttgcattc	tggtagaatt	caactgtaa	15960
tccatctgg	cctggactt	ttttcattag	taggttattt	attactgc	cactttcata	16020
acttgttatt	gatctattca	gggatca	ttcttcctg	ttcagtc	ggagtgtgt	16080
tgcattccagg	aattatcc	tttctttag	attttctat	ttcttgc	agaggtgttt	16140
gttagtattt	ctgttggtt	tttgcattc	tgttagat	gttgcgtt	cctgtttatc	16200
atttttatt	gtgtgtttt	gattttctc	ttattttgc	caaaggt	aaaaagaagc	16260
aataggggaa	ggactctct	ttcaattat	cctactgt	atctgc	ccatatgc	16320
aaaattggaa	ctgttctgt	ttcttaatcc	atatacga	atcaacttac	gatggattaa	16380
agacttaaaat	gtaaaaacc	aaattataa	accctgaa	agaatata	caatatcatt	16440
ctggacatag	gaatggca	agattttat	agaaagac	caaaagca	tacaacaaa	16500
gcaaaaattt	gcaaatgaga	tctaaat	ctaaagag	ctgcacag	aaagaaaacta	16560
ctgtcagat	gaacaggca	ccaaacaga	gggagaaat	ttttcaatc	tatccat	16620
acaagggtt	aacatccaga	atctaca	aacttacaa	atttaca	aaaaaggagc	16680
cccattaaaa	agttggca	gaacatgaa	agacacttcc	cagaagat	tcatgtggcc	16740
aataaacatg	aagaaaagct	caacatcact	gaccattag	gacgtgc	tcaaaatcac	16800
aatgagatac	catctcatgt	cacaatgg	atttataaa	agtcaacaa	catgtatgt	16860
aggttgtaga	gaaataagaa	cgcttttca	ctgttggt	gaatgtca	taattcaacc	16920
actgtggaa	acagtgtgt	gatcc	ggatttagaa	ccagaaat	cattactgca	16980
tatagacca	aagaataga	aatcatct	ttacaaagat	acatgcac	gtatgtttat	17040
tacagcacta	ttcacatag	caaagacat	gaatcaac	aaatgtc	cagtgtataga	17100
ctggaaaaaa	agaatgtgg	acataaac	catgaaat	tatgc	caactact	17160
tgagatctg	tcctttcag	ggacatggat	ggagttgg	gctgttat	tcagcaact	17220
aatgcaggaa	cagaaaacca	accaccat	gttctactt	atagtgg	gctgaaca	17280
agaacacatg	ggcacaggga	gggaaataac	acacactgg	gccagtc	gggtgggggg	17340

tcaagctgag	ggagagcatt	agaaaaaaata	gctaatgcac	tctgggctta	accatttat	17400
gcctagtgtt	ccatttctgg	aatgctaagc	atgtggaaat	tctttatata	ctgctcaagg	17460
tcattgccaa	ggtctgattt	ttcacatca	acaattgcac	acctctggca	taaatgggtt	17520
aatacctagg	tgatgagttt	ataggtgcac	gaaccacca	tggcacatgt	ttatctatgt	17580
aagaacactg	cacatcctac	acatgtaccc	tggaaactaa	aaaatttaaa	atatatatgt	17640
atatatattt	aatatggaaat	tttaaaaatt	actaatagt	tcttttatct	gagtaatttt	17700
gcatcaacat	gcttttatta	tggaaagagaa	gattcagtga	gtacaaaatt	gcagatacat	17760
gtgtcagaag	atcccgtaaat	ataataaaggc	tttagtattct	gtgtcataat	tgccctgttt	17820
tattcctctc	tggtctttaa	acttcattag	ggcaaggatc	aactccatct	tactaaccat	17880
ttgattccct	atgtattaca	cgatataatga	ccaataataa	gcctcaata	aataacttgta	17940
aaataaagaa	tgttatgtaa	tatatacatgt	ggtattgttt	tattgatgtg	ttctttgaag	18000
tgttacctt	gtgtccttag	aggctatggt	gggcctcag	gccatatatg	ttggatgatt	18060
tagtggaaaa	gagattagac	ttttatattt	tattctattt	atttattttat	gttttaatct	18120
tgtgtaaaggc	tatagaggcc	aagatagggc	tcagagttaa	aaacagtgt	gttctaaactt	18180
tcaaaaactac	taaaaacaaa	attagctatc	aaggaaatc	ctgagtgttc	tgactgcagg	18240
ggttccccaa	agcctggagg	accgaagctt	tgaatgctg	catgggtgtc	ctcacctctg	18300
ttgagtatag	tgaggaacat	gagctgagga	gtggattaga	tgaactctta	agatttttt	18360
ccaactttac	ggttgagtt	tatgatgtt	taatgagga	atgagcccg	ggactttgac	18420
tacttgcgt	tgcaactaat	tgctgtttt	ttagctaatt	ctggagtg	tatttgggt	18480
ttaatgttaat	gtcccttcag	cactggaaat	aaacagattt	atgaccctcc	attttcacat	18540
gtcaacttca	gttcccttcag	taaaatgtg	gaagttgctt	ggctgttaacc	actgtacta	18600
tttccctggca	gttttgcatt	tttattacaat	tcgactttt	aatgcattgg	ttcagtccag	18660
ttaataggca	gaaaactata	gctattcctt	ggaactgaca	taggaaagtt	ggtttcttc	18720
tgaggaaaga	tgtgacactc	tgatggattt	ggcaaaaatac	caatttagta	aaatttgctt	18780
ggttccttcc	tgacagcaaa	tggaggttaat	tgtacgtata	tagattctt	cagtaaacaa	18840
agggcaaaaa	tggaaacattt	ttttgttttt	taatgttgct	ggaatataat	atttccatat	18900
ttgggttaatc	atcaaataat	tcaactccat	gcaaagtact	cttttcaaaa	ttacaaatata	18960
tcttgatgga	ttagacttgg	atgtcttaat	tttttattat	gtattttaaa	ttatctataag	19020
gtacattttt	aaaattatata	atacatgtat	cactgtacta	catttggat	aaaatgtgaa	19080
atattaagcg	attaaataact	ttattaatgt	cataaaatttgc	tgccttaccc	aaaatataatt	19140
ttaagaagat	atcatttgc	aataatgtaa	aatattcttc	attattatgt	ttttgttgcc	19200
aaaagtaatg	ttagagtccc	tgttactgt	aattttcatg	actgttaattt	ttcaacattt	19260
ggaaacaaaa	ttttctgtat	tgttttaaaa	tcttaaagaa	tttcagtaga	ttttgacaca	19320
tttttggtaa	aagtatggaa	gtgtcacagc	tttagaagga	tttagaataa	ttatcaatgt	19380
gtatttaggaa	tgacaattttt	aatagacaaa	ttttaactga	aacatttgct	ttattgtatgt	19440
ctaggactaa	aaaaaaatatt	atagtgtatgg	aaccatttca	aaacatctgt	ttttaaaaac	19500
ttctgtcaac	atgtttgaca	gaagtcagaa	taggtaatt	ataaaaaaac	tgccattttc	19560
ttatgtttgc	atcatttagt	tttgcataatc	cttgcatttt	tgcaagtata	ttttaaaacta	19620
cttgttcatt	ttatgtgtgt	tagcctgatt	aattcagata	agatgtcaa	gtcagttaa	19680
gcaggcagat	gtgaaaaaagg	gaatatgtc	tgcccttaact	atcttataagg	gacttactag	19740
cacactgtac	ataacagatg	aacatctgtat	attagatctg	atgggcacac	caatgtttgt	19800
tcataatatta	aataatagtgt	gaagtttca	attaatttag	tttattttt	tatttattta	19860
tttttgagcc	ggtgtttccc	tcttttgcc	aaggctgggg	tgcaatggca	catgggtgca	19920
gctcaactgca	atctccacat	cccagggtca	aaggattttt	ctgccttac	ctcccaagta	19980
gctgggatta	caggcggtgt	ccactatgcc	aggctaaattt	tgttattttt	agtagagaca	20040
agggttcacc	atgttggta	ggctgctctc	aaaatcccg	cttcaggtga	tccacccggcc	20100
tcagcctctt	aaagtgttgg	gattacaggc	atagccact	gcccctggcc	aattgtat	20160
ttatgtttaga	aatatataat	ggagttacaa	tactttgttt	ccaatgttcc	gctggaaacc	20220
actgggttggaa	aggataggtc	ccagtgctc	ctgcccagtt	ggcagttttt	gatagaataat	20280
aagttctatg	aaggcatgga	tgaagtatcc	ttcatggtaa	tgtatgttaat	gttagtataag	20340
tcacttggtt	aatgttgtat	aactttttaa	ctcttccaac	ttcaatttca	tctatatctt	20400
tcatagcagt	ccagtaaagt	agtatcaatt	atagatatca	ttcctacttt	cagataaaagt	20460
aactgaggtc	cactagttt	gacttgc	agtttacgt	aatgataagt	gacagactgg	20520
atttaaaggg	cgaggatccac	tttactaaac	tgccctctgtc	ttctagccaa	gtttgttat	20580
taaagtaagt	attcagacat	tggctaaat	tatttataat	caagaaaaca	gtagaacagg	20640
ttaaggaat	tttttttcc	ctctccctcag	tgtcaaaaaag	cagtcagta	aaagatccaa	20700
gataaataac	ttcagttccc	aaaaataact	tcaattccaa	aatttccctt	tgacttaggt	20760
tatctaattt	gatagtctat	aaaatgtgt	ggaaattgaa	gaaggat	tttggat	20820
tgtggcaact	taatggtaat	attgaagaag	caaattggata	tttattgtat	catataatgt	20880
ttttaaactt	aaatactagg	gtttgtat	catatgtttag	gaaattatct	atttgttgca	20940
gtagaaat	gcataaatga	atggaaactt	aaattatctc	atctgtgca	ggcagtc	21000
agagactgt	gtttatataa	gtgaaaactc	agaaaaattt	tttgcgtc	tttcaagctg	21060
gatagcgggg	tcaaaaactc	aacttctaa	tttgcgtat	aacgtcagaa	atcacaaaag	21120
aagtatttca	cagaaaaaaa	tgacatacc	cggaaaacatt	caaatacaaga	aaatcacagg	21180
ggtacagtgg	atggcccccc	ttaaggtaat	gcgtacttgg	cttcccccgt	ggttaatgcc	21240

tctctttac	taagttgttg	tcgtgcagag	tggcattgca	gctctgtact	tttacagaat	21300
aatttcagg	tctaaaattt	atacattatt	tcagattgt	agaatagtat	aaaataaaaat	21360
tttaaaaggg	aaaataattc	catcaatta	atggaatgt	tagggtttaa	gttataacaa	21420
caaaaacaaa	gtttaggttt	tttaggatta	ctcatataaa	tagtgtgtt	attaagaatt	21480
actggcttt	tgaatattaa	ataagaaggc	tggccgtgt	ggatcatgcc	tgtaatcccc	21540
gcactttggg	aggccgaggc	agacggatca	cctgaggtca	ggagttcaag	accagcctgg	21600
ccaacatggc	gaaaccctgt	ctctactaaa	aataaaaaaa	ttagccggc	gtgggtacgc	21660
atgcctgtaa	tcccactca	tcagggcgct	gaggcaggag	aattgttga	acccggaaga	21720
cggaggttgc	agttagctga	gatcgccca	ctgcactcca	gcctgggtga	cagaacaaga	21780
cfcacatctt	aaaaaaaaaa	aaaattaaga	caaccacaga	tataaggatt	tcaattaga	21840
aagcttgaag	ggaaattttt	ggaacttcta	ggtcaaccta	aaattacaaa	taggaaagca	21900
gctttggaga	ggaacatttg	cacttctgaa	gtaatactgt	gatttggact	gtcagacatt	21960
gtgatataatt	ggcatcacca	cacctctgtt	aattataaac	ttacctattc	tcagtgcctt	22020
gttaactgtt	gaaataggac	ttgacattt	ttttcttcc	tcttctgtt	ttcattttt	22080
cactcattt	gttagtaatt	acttcttagt	gccagatagt	tttctaggt	ctgaaataaac	22140
aaaaatccta	ttgtcatagt	tatgtttt	tggaaagaac	aacctgatag	actatcgta	22200
aaaataaaatg	agttgtata	cataaaacac	agccatgag	atatagtctg	taattatcat	22260
ccctgcttca	tttattttt	tatttattt	tttattttaga	gaatggatct	gattctgtca	22320
tccaggccgg	agttcagtgg	ctggatcatt	gctcaactgt	acctaaaaac	acctggcctc	22380
aagtgattcc	cccaacctcg	gogttccaaa	gtgctggat	taaaggcatg	agccactgt	22440
cctggccatt	getctgtt	taaagatgag	gaaactgaaa	gtacagagag	gaacgtgact	22500
tgctcaggat	cacacagcca	atcagtggca	gagcagtcta	ggcagtcgt	gcctgaaggc	22560
attattctt	ctttctgttc	tgctcaaaa	accctaggca	ggcatggtaa	aaagactgaa	22620
agagccaaagg	caacataggt	acctgaagaa	tgggagaaca	cctatgagga	ctgctagaaa	22680
ttttaggcag	ccctttaaag	gccttggca	acaatttga	cgtgactcat	gagtcaagt	22740
attctctgca	ccttttttt	tttagagaat	acaatggata	aagcaatgac	ctttgttaaca	22800
tgtaaacctg	tatttgaatc	ctttccttc	catttaattt	tgcagagtca	ccttttcttc	22860
attagaaaga	aattttatcc	caaagagatt	atcgtgaggc	tcaaataata	taggtgaaag	22920
cccttgggt	ttaaaaaaaa	gcctttctg	tcatttctgg	gtatctccg	tcagtgatt	22980
tagtaattgt	ttacctctct	ggcttctgt	gagaagacaa	attgttagaa	gaagtggagg	23040
atggcattgg	gttaggtgt	tagttctgt	tcacccactt	ccttaggtct	ttttttgtc	23100
ccccaaagact	atatcatttc	tgttctct	tgaacctgt	agattttatg	gtaatctgt	23160
caatatttat	aaagtatgaa	atctctgca	tttggcaca	aaagtggca	tagacaatgt	23220
ataaaatgaat	gggcattgt	ttctccagta	aaactatatt	tatagaaaaata	ggcagtcctc	23280
tccaatttt	gtttgtc	cccttctca	tagaactcag	gtttatcata	tacaatgtta	23340
cacattattt	ctttgaataa	attgttacc	taatactttt	taaatttggg	cattttctag	23400
gacattctat	ttttttt	ttaacttgt	gttatactgc	agtgaatgtc	ttccctgtgt	23460
tagcatttt	tttgggtatg	tgtgtcaata	tctataaagt	agctcttaag	aactctggaa	23520
acattgtca	actgttctt	caaacagcag	agggtggagg	gagcatcgcc	ctaaccctac	23580
ctggcctgcc	ttaccagcat	gcaaccaagg	agaagagaat	aggatcaaga	gccagtgc	23640
gagagaagta	cataatacaa	ctttatggag	tttaggttgc	atcctataaa	gttaccatgt	23700
tacactgtt	cgatctacca	aagtagaagt	tttacatagt	ttagtagt	gtaatttttta	23760
ttaacctttt	cacacaacac	accctctcat	attgctgccc	aacaatgtcc	ttgggttctt	23820
tttccatgtt	catcatattt	agtaggttt	tagcagggac	ttgagtagt	ggaaactgaa	23880
ctcatgctag	tgaagccatt	cttcactaga	agcagacata	tgggatccag	aatatcatgt	23940
acaaccagct	gggtacagt	gctcacac	gtaatcccag	cacttggga	gtctgaagg	24000
agcagatcac	ctcaggt	gagtttggaa	ccagcttgc	aaacatgt	aaaccctatc	24060
tctactaaa	atacaaaaat	tagccaggcc	tggtagcaca	tgcctgt	cgccagctact	24120
tgggaggctg	aggcaggaga	atcgttgc	cctgggaggc	ggaggttgca	gtgagccgac	24180
atctcgccag	tgtactccag	cctggcaac	agagtggac	tctgtctgaa	aaaaagagaa	24240
tatcatgaat	atcaccaaag	ataaaagcaa	gatgcggata	tgagtagccc	agagtaactt	24300
ttcctgtct	atataacaca	gtgagttgt	attcccacaa	ttcaagtcca	gtttttctt	24360
attctaaagc	catacatttt	taactctccc	cctaaatatg	atttgaattc	agagagtca	24420
atgcaataag	gtatattttaa	ctttaagtgt	agtttagccc	acacatttct	ctaaactggg	24480
ggttggcaaa	ctacaaacct	ctttttgtaa	gtttaagttt	attggaacac	agcctagctc	24540
attcatttt	gtattccctt	tgcaccttt	agtttgagta	gttgtact	aggtcttca	24600
gccttca	ctaaaatatt	tactatctag	tccttcaaaa	agaagttgc	caaccctctc	24660
aaacaagtg	aataatgt	attagatgt	agcaggagtc	cacaaactac	aacctctgga	24720
gccaaaacca	ccctactgtc	tgtttttgt	aacagaaattt	ttatagaaaac	atagttgacc	24780
cctgtgcca	tttgggtt	tctgtttt	gatgggtttt	gtgctacaag	ggcagagtt	24840
aacagttgc	agagaactaa	gtggcc	aaggctagtg	gtctttgt	tctggccatt	24900
ataggaagtg	tttaccaa	cctgcattt	atcatggat	cccagaatgt	ttatacctt	24960
gaggattttt	ttttctactt	tccatatttt	acaggcaaca	gaactaaagc	ctcgaaaat	25020
tttgc	tttgc	tagtctt	tattaaatta	agtcagagcc	tgaattgtc	25080
tttgc	tttgc	gaaatttagc	acagcatttt	cccttctgt	ctcccacatg	25140

ctacacagtc atcctggatc atgtccccg cctccatgag gagatgttct aacatgaggg	25200
tgactaggct ctggagccag agcatcttag ttcaaattctt gatatcttta aaaataataag	25260
ttattttca gttctaacac tgaataaga caaaaggcca tattactgaa ataaagtgtt	25320
aagaatataat ttaatactca gttactta gcccctcat gatcagaata tgtaataaaa	25380
ttgagagatt gacagaaaaa aaggattctg tttactgtg gtcacac gtaatcccat	25440
cactttggga ggccgaggtg ggtggatcac ttgtctggag ttcaagacc gcctggccaa	25500
catggcaaaa cccctctct actaaacaat acaaaaat gccaggctg gtggtaagca	25560
cctttaatcc cactactcg ggaggtttag gcaagagaat cgcttgaacc tgggaggtgc	25620
aagttgcagt gagccaagat cacaccactg cactccagcc tgggcaacac agcaagactc	25680
catccctgga aaaaaaaaaa tcatcatttgc ttctttagt ttcttagtgt ggcccttgc	25740
tcaagtggaaa attttagtgt gatagactga ggagaattct gtttctgcca cttgcctgt	25800
tgccttggta ggagctcctt atctctgagc tgagctcctt ttgtccttgg taaaatgtgg	25860
ttattatcat ctgcattctc atagtcttct tgtaaggact gtgaagtac tttgaaaaaa	25920
tgccttgaat atgcagatg gtctgtttt tttttttttt gttttgtttt gttttgtttt	25980
cccttggcct tcattctgg aagtgtttt ttatatacat tttgtttggt ggccaacagt	26040
gctaaccaca gaaaagttt tatgtcttct tttcttagaa taattcttac agtgaacact	26100
ctcagggtga agttttgtt ttgaaagggg cacattact gtgaatgaga gccatggaaa	26160
tggacctgaa atttgaagac agatttact tggttttctt ttcttcttatt tttttgcagt	26220
ataaagaaga gggcagcaag gtgaccactt actgcaacga gacaatgact ggggggggc	26280
atgtatgtttt gggccggaa tgggcttgg tcaaccggaa gaaggtggaa actgcctctg	26340
agaatgtgt tgcacacaca gcacacccat tttttttttt gttttttttt gttttttttt	26400
atgagacata ctgtttttttt tttttttttt tttttttttt tttttttttt tttttttttt	26460
attttagttt gattataaa aatatcttgc gcttagaaat attaggcatg ctatgtaaaa	26520
ccttactgaa aaaataattt gaccaacattt attttagtgc tttttttttt gttttttttt	26580
agtaatggac caatgattttt tttttttttt tttttttttt tttttttttt tttttttttt	26640
actaataaaag aaaagccatt ttgtttttttt tttttttttt tttttttttt tttttttttt	26700
agatacgtat gtttatgttca caacaagtat aggtgataact aattggggaa cttttttttt	26760
agaccataaa taactgaaat caaatcttca agtattttatg cagcctgttgc gtttgcgagg	26820
gctgccataa caaagtacca cagactgggt ggcttaagga aggttgggg ttcacagttt	26880
tggaggctag aagtccaaaga ttgtttttttt tttttttttt tttttttttt tttttttttt	26940
ccttggcttg cagatgtctg ttgtttttttt tttttttttt tttttttttt tttttttttt	27000
taatgtccctc ttcttataag ggcattttttt tttttttttt tttttttttt tttttttttt	27060
aatttttacct taatgaccc tttttttttt tttttttttt tttttttttt tttttttttt	27120
acaggggggt agaatttcaa cgtatgtttt tttttttttt tttttttttt tttttttttt	27180
gcagttaatg tttactgttta aattttttttt tttttttttt tttttttttt tttttttttt	27240
ttcaaaattttt tagtattttt tttttttttt tttttttttt tttttttttt tttttttttt	27300
ttatgttagt gtaaaaagcca caaatcttca agtattttttt tttttttttt tttttttttt	27360
attagtaatg aatgactgtca gtcattttttt tttttttttt tttttttttt tttttttttt	27420
tctgcaaaagg tgctgggcaaa gatctttttt tttttttttt tttttttttt tttttttttt	27480
ttaaatatgt ttgtttggct tttttttttt tttttttttt tttttttttt tttttttttt	27540
tgtaccctat catttaggtt tttttttttt tttttttttt tttttttttt tttttttttt	27600
tctctcattt agtgaagaag ctcccaaccg gaggctttttt tttttttttt tttttttttt	27660
tgaagtggc atgtggactt gttttttttt tttttttttt tttttttttt tttttttttt	27720
tatcccatag taagatacat tttttttttt tttttttttt tttttttttt tttttttttt	27780
gtatTTTAAAGA CGTGGGATC TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT	27840
tttttgcataa tgtagatgc tttttttttt tttttttttt tttttttttt tttttttttt	27900
caggcaatgtttt cttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	27960
ttttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	28020
ttttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	28080
ttttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	28140
ttttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	28200
ttttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	28260
ttttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	28320
ttttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	28380
ttttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	28440
ttttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	28500
ttttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	28560
ttttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	28620
ttttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	28680
ttttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	28740
ttttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	28800
ttttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	28860
ttttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	28920
ttttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	28980
ttttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	29040

aaattgatta	tgattatgt	taaacttga	tatcttctca	tatttaatat	aatatagtt	29100
tgagcact	tccatcat	taaataactt	tctagacatt	ttaaatgaat	atataaaaaa	29160
tgaacatca	ttaatcagtc	ccttatcta	gaatatttat	ttcggttatt	tacaatctat	29220
tagtagttt	aaagacttc	atggaaaatc	cttatacatg	catcttagtt	aacttgaact	29280
tttattttt	gagaagacat	gccaggctgt	tctaggctat	tgatacatat	tgccaattac	29340
cttccctgaaa	gattgtacca	cttccactt	aggcacagag	tgtgaatgcc	tgttttaatt	29400
actctattt	catcagtaaa	gactatatct	gaaattatct	cattttggg	gacaggtatt	29460
ctaatacggt	ctacaagtt	gatcacaact	ttgtgaaagc	tatcaatgcc	attcagaagt	29520
cttggactgc	aactacatac	atggaatatg	agactcttac	cctgggagat	atgatttagga	29580
gaagtgggt	ccacagtcga	aaaatccaa	ggtaatcaag	cacacatttt	atcattaata	29640
aaaatatgaa	tgctgaatac	catcctccct	cctaaagcagt	cctaccagt	ttgctcgcca	29700
ttttatttgt	catcccagtt	gtaactttt	atgtgatctg	ttaacgatct	ttatctccta	29760
ttgatacata	ttctgtca	ctccaagttc	tatttatgag	tttttcttc	acatgtctgt	29820
tgtttccatt	gcccccttc	cattttctac	actctggcat	gcagcttgc	cctgtttccc	29880
ccatgctatg	gattgcctc	ctgcccgtaa	ccacctgagg	ggttcttct	gcctgctgca	29940
taaagaaaaaa	ccatggcaat	atagtagaga	aagagcttaa	gagacactag	gttggccatg	30000
ccacgtgggg	gatggagttc	atattcaat	tatcttgc	aaagctata	gttaggggtt	30060
tttcaaaggc	agcttggggg	aagggggtggg	ggtggccagg	taacagatgc	ttgctgtga	30120
ttgggtgggg	tgggtgaat	cacagggt	tgaagctgtc	ctcctgtggg	ctgaaattgt	30180
tctaggtggg	gccccatggag	tggggttgtc	gggtccaggt	agaaccacgg	gtgtcagaca	30240
tgcaaaaaata	aaataaaaaata	aaataagata	angttaagat	aaaataaaaat	aaaataaaaat	30300
aataaaataa	ataaaataaa	aatttccctg	aaaagatatt	tcaaaaagcc	agtcttagat	30360
tctacaataa	tgatgttatt	tgctggagta	attgatggag	ttgcatgtct	tataacctct	30420
ggaataaacgg	ctgacaatct	ctcaagtc	cgcccttagct	ggactcaggt	tcctcttctc	30480
cccacagcct	gactgcctcc	attagctca	caaaaagtgg	tgggttcc	ggcaaggccc	30540
attgtcattt	aaactgtac	cgaaatgact	tccaaagtt	gttggccca	atagcccagg	30600
aatatttaag	tggaaggca	gatggggat	gggttagctt	agctctctt	cactctcata	30660
gttttctcac	tgttataatt	tttgc	aaagg	cggttcatg	cctgcacatct	30720
tacctctccc	agttcccatt	cttagctgtt	ttatgaaatg	cttctagttt	catcctctta	30780
taccaagttc	tgggagactg	atttgagtaa	taataaaact	ccagttccc	atacagccgg	30840
ctctgcgtg	atttaaactca	ttttctattt	caatttccct	gttgcatttt	tcagttctgt	30900
gttaggcgtg	aggaaggaga	acccttggg	tgattacgag	actgtgttac	tgcccaactac	30960
ctaatggata	catttagct	ggtatccaa	cccatctaat	tatgaccata	actatattta	31020
tcaccttgc	ctgccttcaa	acgatatgac	acacaatgaa	tgaaaacttt	catttttcat	31080
cttcattttgt	gctgttccct	ttgcctcaa	tagccctaa	ttgccttaca	gtactgtaa	31140
aatttgcac	ctaaaaaaaaaa	atctcaaaat	cctctctatg	cttgcatttt	cagaaaaaaa	31200
aaaaaaaaaa	aaaaaaaaagt	tatcatttttata	gtcttattac	tctcaaccag	aggtggcatt	31260
tttcatctt	accataa	cccatggat	atttcttagt	ttggaaattt	atatacataat	31320
actttgaatc	tgtcttagtca	gaaattttaga	taatttaaaa	aaaaattattt	tttctgaaag	31380
tatagcttga	tctagcttag	gatctacatc	tatctaata	agttccctaa	catnttgtta	31440
aatggccacn	gnnataagtc	ctgtatgcc	atactttgt	ttaggatcat	gtgactaagg	31500
ggttaaggaat	tggaaagca	tggggagct	gcagaattt	aaaaattata	tagtaggtt	31560
atttttccctt	aatacattga	aatagccca	aattctcaga	gaatacaatg	ttaatecc	31620
taatatactgt	agagttcatg	gctcaaattt	gttattttt	aataaaactac	taatagatta	31680
ataattactc	aactaaaaca	cttgc	aaatgg	tgaggttcc	cttacccccc	31740
atgacagagg	tgtggcttc	cacaccttgc	gttgc	ccctaccacc	caaaccctta	31800
ggggggagcat	gcagggggc	aggctcagag	gcacatgggg	gtgttttgg	gtctggccc	31860
cacacgtg	tctaggatgt	gatgttggag	actcctgaa	cccaagtggg	catgtgttac	31920
agtgtgtct	ttcagcttag	ccgtctgc	atggcttgc	ttaatcagg	cattagaccc	31980
catgccttat	tgcaagggca	ggggtccaa	gtgacagct	aagttctgc	tcagtgta	32040
agaagaattt	gatcacacgt	gggcttggaa	gatgagcaca	aggtttattt	gagtgggtt	32100
ggtggcttc	cgcgagacga	ctctcagcca	gagaaggagg	atggagggga	aagggtttt	32160
tccctggag	tccacttgc	cctgttcc	tctgggttca	aataccctt	ccctccctt	32220
ctgtgtgt	gtccaccac	tctccacagc	tctgtccgc	tctgttcc	tgctcctct	32280
gatgttgc	tccttgc	tgtgc	ttttat	gggaggatg	32340	
ggggggatgg	cgggccaaaa	ggcacctt	ttgggtgt	aacagaaaatg	cctgttctt	32400
cttagggcct	taggtcttca	ggcttggaggg	tggggcctt	gctgagggac	caccctcttc	32460
tacccagat	ttccctgtct	cctgttccata	tcaacattac	taactttt	atctgcagac	32520
taataatgt	aagggtgtgc	attttcaac	tgtgagacta	tgtgaaggtt	tttctgttcc	32580
aactgtatgc	atcccccatt	aatttctacc	cttctttaaa	agaatcttt	gcagtttt	32640
tccaa	ttcttagagaa	tttcttgc	gtgaaatgt	ctagtaatc	aaaattggaa	32700
aacggagcat	atcatatccc	cttctcaaa	tcaacaa	gaagtcctaa	tgtgtctt	32760
tgtatctgca	tgagacagga	agctgagatc	tattcaacaa	caaaaatcca	aacaagcatc	32820
aagaggagga	gtgttagcac	ttgagctag	ggagactgt	gctcctgc	gaaagatgg	32880
agcctcagtc	acagctgctt	taccaagt	catatgc	tttctgagg	actcctgc	32940

aagctccctt	ctccctccag	ccaaccactt	ttgttttaga	caagggctgg	gtttatgaag	33000
gactgttttc	atgactaaag	ctttatagaa	ggttaagat	aaggagatgg	aattgagtga	33060
agtaggaaat	atgaaagcag	atattataat	ctggttccct	gatttttcac	tagcattttt	33120
gtttataaaat	tagtctgtt	ctaagaatcc	aatgacgtaa	tagaaactct	caaagattct	33180
taacttgaga	tatagggagt	ctttgaaact	gctgaaatta	cagacagcat	ttattgttta	33240
tgagcatttc	tgaatctaga	gcttcacta	gattgtaaa	gaatgtggc	caaaagatta	33300
agagccaaatc	ctgtatcttg	tactcaaaat	gttgttaatt	ctcaccttt	tatcccccaag	33360
actctcttg	ctctetcttc	attttcacag	tgtaactgg	aagggtcct	gatccagacc	33420
ccaagagagg	gctttggat	cttgtcaag	aaagaattta	ggcgagatcc	tcaatgcaaa	33480
gtgaaagcaa	gtttattaag	aaagtaaagg	aataaaagaa	tggctactcc	atagacagag	33540
cagccctgag	ggcagctgg	tgtccattt	tatggttatt	tcttgattgt	atgctaaaca	33600
agggatggat	tattcatgcc	tcccctttt	agaccataga	ggtaacttc	ctgacattgt	33660
catggcattc	ttttttttt	ttgacagagt	ctcaactgtgt	tgtccaggct	ggagtgcgg	33720
aaagcaatct	tggctcaact	caacctctgc	ctctggggtt	taaatgattc	tcctgcctca	33780
gcctccaaag	tagctggat	tacaggcatg	accaccatg	cccagccaaat	ttttgtattt	33840
ttagttgaaa	cagggtttt	ccatgttgc	caggtggc	ttgaactct	gacttcagg	33900
gatctgccta	ccttggcctc	ccaaagtgc	gcaattacag	atgtgagcca	ccaaacctgg	33960
cctgtcatgg	cattctaaa	ctgtcatgtt	gctgggtggg	gtgttagcgt	gaggaagacc	34020
agaggtca	ctcatgc	tcttgggtt	ggtgggtttt	agctggctc	tttactgcag	34080
cctgtctgtt	ctatcagca	ggtctttatg	acctgaatct	tgtgctgacc	tcctatctca	34140
ttctatgact	tagatgcct	taactgtctg	ggaatgcago	ccagtagtc	tcagcctcat	34200
cttacccagc	tcctattca	gatggagtt	ctctgggtca	aacaccttt	acaacatcat	34260
taagcctcag	ttctcacact	gtttttgtt	tgtttgtgt	tgcaatgatg	ggcaaatctc	34320
tgccctata	gatgttagaa	aaaaggaact	taatattgt	atgactgttt	tgtccagat	34380
aattgcttta	aactgtatca	geatcttatt	tagtctgtt	aatgatatga	atgttatctt	34440
cataacagct	gccattttat	taaggactt	tcagagaaaa	acactgttct	aagcacttgt	34500
tacccattat	tacattgaat	tttcataaca	accctttgag	gtaagcatga	ttataccac	34560
ttaatagaa	gtgaactgt	gtttttaat	gttaggttcc	ttgccaaatg	ttacacagat	34620
agtaagtgt	aaaatcatat	gcccgtaaat	tacattatgc	tgccaaactt	aaatttctt	34680
tttacccattt	atattagat	attcttagt	ttaaacaaga	caacttagtt	acacatacta	34740
gattttgtcc	acagttctg	gctcataact	ccatagccc	ttgtcaat	ctttaancg	34800
ttggggcaca	ttaggtctc	gaagtagggc	tcangaaaac	aggtctctc	tctctctc	34860
tctctgtat	tctccgccc	tcctctcacc	tgcccaggc	agcaattaa	tctctctctg	34920
ccttctgtat	cttgggtcat	aagacctca	tttccaaaga	tgtcctgtgt	cataccctaa	34980
aggaaggaaac	actgaacaga	gagaggctca	gaagaatctg	gacaggcctt	gctgtgtt	35040
catcattccc	tttatgtca	gtcacatctc	tacatggtt	tcagttgtc	ctatttgatg	35100
aagtccccat	ataaggctca	caaggacagg	gtcagagag	cttccagata	gctgaacaag	35160
tggaagttcc	tggaggggtgg	cgtgttcagg	gagggcatgg	aagctgtgt	cccttcccc	35220
cataccttgc	cctactcatt	tcttcatctg	tttcattttgt	agtatctttt	ataataaaacc	35280
actaaacatt	agtttagtatt	tctctgtatt	ctgtgagtc	ctctagcaaa	ttaattgaac	35340
ccaaggaggg	tgtcatagga	tcccnacat	tatagctgg	tggccagaag	cacaggtaaa	35400
caacctaggg	cttcaattt	gcatgagaag	tagggggcag	ttttgtggg	cgagccctc	35460
agcctgttag	atctgtatcc	atctctaagt	acacagtgtc	aaaactggat	tggaggacac	35520
ccagcttagt	ttcactgt	aattgggtc	ttgcttatt	tgtggggaaa	aaccacatg	35580
catttgcata	cagaagtctt	ttgtgttgac	agttgatagt	gttcagtgt	agaattaaaa	35640
aaaaatttag	tttctcttc	aacatactct	ctcaatgt	aaccacagaa	actatttcca	35700
ttcaaaatgt	gaaatggttt	gtttgcattt	tagttttat	ttatacatct	tagaagaaaat	35760
gtccaaatgt	tgtttttttt	ctcaccctat	atataaaatt	acctatgagg	cacagatttt	35820
tatgtccat	gattatata	actttgttca	aattgtgtgt	ttttagatcat	tactgttaact	35880
tgttata	atctttgtgt	atattatgtc	tcttaacatt	gtcttccata	ttgttaatga	35940
ccatctcata	tttatctctg	tatcatgtat	atctcaacc	aatgtgactg	gcttaggaga	36000
aaaaatttagt	gaacaattaa	ctagtgttt	tgtatctat	acaattgtca	aggttacaat	36060
tgctat	ttttttttt	ttgttggttt	tctctttgtt	tcatctcagt	tccattttgt	36120
caaggattcc	ttttttttt	tttttttttt	tttttgaggc	ggagtcctgc	tctgtcaccc	36180
gggctggagt	gcagtgg	aatctcggt	cgctgcaagc	tccaccttct	gggttcatgc	36240
cattctctg	cctcagc	ccgagtagct	gggactacag	gcacccgcca	ccacgcccag	36300
ctaattttt	tttgat	tagtagggac	gggtttcac	tgtgttagcc	aggatggct	36360
caatctctg	acctcg	ccgcccgc	cgccctccca	atgctggat	tacaggcgt	36420
agccaccgc	cgtggcc	tgtcatgtat	tcttaac	tgttatatct	aagagaaggt	36480
gtgagaggcg	ggaccattt	tggataggt	cagagggcat	gaagttagccc	agaaaagaatc	36540
tttgcattt	actaaat	agcttagt	aaaacatgt	gtaaccactg	aaaaactaga	36600
acagtgtat	atgacaat	ccagtgt	aagaaattca	aatagttat	agtaacaaa	36660
aataattttt	accacattt	gacttagt	ctaaaaacaa	atttgaagga	gcagagagag	36720
agaggaagt	agtttcgt	ggggctgt	gagtggaggg	tactttatga	gcagtttga	36780
aaattacaag	tttagggaaa	actcttaggg	gagtttaag	tggtagca	caggcacttg	36840

aggttcagg agcagcagca taaagatata gaacagagag gcccatttggc agtcttggg 36900
aaactccaag aagtcaatg tgtctaaagc tgtggtttc tctctgtgt tgcgtgtg 36960
tgtgtgtgt cgccgtgtg ttgtgtgtg tgtgcgtgt tatgtatgt tgaaatcgct 37020
aatacagaga agctgttaga acaattttct cacttccaa aaatgtccgt tcattcatgg 37080
gcctagcaa ctctccctt tgcgtttac agccaaacac tgcataataa ggtgtggtc 37140
tataatgaaa cagtttattc tgcgtttac ggagctagtc agcctgtgtt atgcgttta 37200
ctgggttggagg gtgtccagggt tgcgtttcc ttgaacaaag aattggacaa aactcacaaa 37260
tgaggcaagg aaagaatgaa gcaacaaaag cggagattt ttgaaaatgaa aagcacactc 37320
cacaggatgg gagtggccct aagtaaatgaa ttcaaggccc tgggtacaga atttctggg 37380
gtttcaatac cgcttagagg ttcccatgg tgcgtttcc ttgcgtgtt aatggagagg 37440
gagggtggag ttaagtaca aagtcattt ctcgtatag gcctgtgtt aatggagagg 37500
gtgttactcc tgggggttggt ggcccatgtaa acggagagg atgaatgtaa gtgacaaagc 37560
ccttcgcatt ctgcatttgc ctgaaatgtt tccactttat ttgcgtgtt tgcgtgtt 37620
tgaattggcc ttatgttccc tgcctccaga acctgttcc tgcgttccatc tagaccctt 37680
tttcccttgc ccaacttcac ttcttttca gcccactc tccaaatgtt tgcgtgtt 37740
aagtcaaaaca tcagtatttgc ctctgttttca tctaaaccct tcttatgttgc tctacttctc 37800
atctcttgc ttgtcaaaaaa ttcttcaaaa atgagtaaga atgcgttat tcaggcaaac 37860
taaaaataac atcacatgt tatacaaaaac cagtgccatt tcacaaagga aaattatcaa 37920
taactatgtcc tgaaaaaagaa acagcgaatg aaagccattt acacaactcc attgtgtat 37980
tgacacatttgc aatcactcat aaaacaggtt ctctgggtctt gaatctagat cctagctgt 38040
ctggtagctg aaatcataga attatgttagt agtttagaa atcatcctca aaggaaagat 38100
tatatgttgc tatcaatgt atatttccct tctaggccca aacctgcacc actgactgt 38160
gaaatacagc aaaagatttt gcattgcca acatcttggg actggagaaa tgcgtgtt 38220
atcaatttttgc tgcgtgttgc tgcgtgttgc ggtaaaaaaa taagcctaag tttttgtt 38280
atttgttgg aactatttttgc tgcgtgttgc tgcgtgttgc tggatttccgg ggatacttag 38340
atggaaatggg catgatcccc tttttacaga aatagaaaat agtggcccta tgaattattc 38400
tttccttttgc tatccatgac aaacttttagt aaaaaatttt ctgttctact gagcttagca 38460
tttatttgcactcttgc atatatttttgc caggttagctt gtgacaatca gagtgtatg 38520
taagacaaac tcatttgc tgcgtgttgc aagatttca gtagaagcaa agaatttgc 38580
accatttgc aagtgggttgc ttggaaattttgc tgcgtgttgc tggatttccgg ttttttttgc 38640
ccccattatatttgc ttagtgcatttgc attactatcat tgcgtgttgc tggatttccgg 38700
tgcgtgttgc ttttttgc tgcgtgttgc tgcgtgttgc tggatttccgg ttttttttgc 38760
tatgtgttgc taagtattaa cccagcttgc tgcgtgttgc tggatttccgg ttttttttgc 38820
tcatgaaaatgc tatttgcatttgc attgaatttttgc tgcgtgttgc tggatttccgg 38880
tgagttgatt tcataacatgc tgcgtgttgc tggatttccgg ttttttttgc tggatttccgg 38940
gagttgtgc caacaatgac ttggagaatc tgcgtgttgc tggatttccgg ttttttttgc 39000
ctcattttgc agataataact acccatggct tgcgtgttgc tggatttccgg ttttttttgc 39060
tgcgtgttgc aaacatttgc attagtgct tgcgtgttgc tggatttccgg ttttttttgc 39120
aagantaaac ntgttgc tgcgtgttgc tggatttccgg ttttttttgc tggatttccgg 39180
gtctcattttgc tgcgtgttgc tgcgtgttgc tggatttccgg ttttttttgc tggatttccgg 39240
gaatgttgc tgcgtgttgc tggatttccgg ttttttttgc tggatttccgg ttttttttgc 39300
atctgacacc aagggtctgc tgcgtgttgc tggatttccgg ttttttttgc tggatttccgg 39360
ctaaagaaatgc tgcgtgttgc tggatttccgg ttttttttgc tggatttccgg ttttttttgc 39420
ctttggaaatgc tgcgtgttgc tggatttccgg ttttttttgc tggatttccgg ttttttttgc 39480
gtgaaatgttgc tggatttccgg ttttttttgc tggatttccgg ttttttttgc tggatttccgg 39540
tgcgtgttgc tggatttccgg ttttttttgc tggatttccgg ttttttttgc tggatttccgg 39600
ggcacatgttgc tggatttccgg ttttttttgc tggatttccgg ttttttttgc tggatttccgg 39660
gtgttgc tggatttccgg ttttttttgc tggatttccgg ttttttttgc tggatttccgg 39720
actaatatgttgc tgcgtgttgc tggatttccgg ttttttttgc tggatttccgg ttttttttgc 39780
cattaaatgttgc tgcgtgttgc tggatttccgg ttttttttgc tggatttccgg ttttttttgc 39840
aagagaaggttgc tgcgtgttgc tggatttccgg ttttttttgc tggatttccgg ttttttttgc 39900
agataataatgc tgcgtgttgc tggatttccgg ttttttttgc tggatttccgg ttttttttgc 39960
ttcataatgc tgcgtgttgc tggatttccgg ttttttttgc tggatttccgg ttttttttgc 40020
aagttatgttgc tgcgtgttgc tggatttccgg ttttttttgc tggatttccgg ttttttttgc 40080
ctttgaatttgc tgcgtgttgc tggatttccgg ttttttttgc tggatttccgg ttttttttgc 40140
aagtttgc tgcgtgttgc tggatttccgg ttttttttgc tggatttccgg ttttttttgc 40200
ttctcttgc tgcgtgttgc tggatttccgg ttttttttgc tggatttccgg ttttttttgc 40260
ctgtataacc tgcgtgttgc tggatttccgg ttttttttgc tggatttccgg ttttttttgc 40320
cagagagcaatgc tgcgtgttgc tggatttccgg ttttttttgc tggatttccgg ttttttttgc 40380
tttacacttgc tgcgtgttgc tggatttccgg ttttttttgc tggatttccgg ttttttttgc 40440
gtatgttgc tgcgtgttgc tggatttccgg ttttttttgc tggatttccgg ttttttttgc 40500
agctatgttgc tgcgtgttgc tggatttccgg ttttttttgc tggatttccgg ttttttttgc 40560
gtcacattaa tgcgtgttgc tggatttccgg ttttttttgc tggatttccgg ttttttttgc 40620
tgcgtgttgc tgcgtgttgc tggatttccgg ttttttttgc tggatttccgg ttttttttgc 40680
caccatgc tgcgtgttgc tggatttccgg ttttttttgc tggatttccgg ttttttttgc 40740

ccccgtgttg	cacaactaat	gataggcaaa	acaagagagg	agggaaaaaa	aaagaatcc	40800
gactcccac	tcctagttca	gtgttctt	cactgttaga	aggtgcgtct	gaacatagt	40860
ataccatatg	aagatcacac	tatctattt	agattgtaga	aatttggatta	ctgcagagct	40920
ctggctggc	tagattca	tcttatttt	tattgcattt	cagtatttt	aagagataaa	40980
tggctcttt	agaatcagat	tggccttgc	tgttggaaat	aggcatataat	taccttggta	41040
gtgacacat	tcttacaggt	caggggcttgc	atgaagttt	tcttcttccc	tttttctgt	41100
gattgtctg	taaatagaac	acattcagag	cccttgaat	cacttgacgc	ctgtgcctcc	41160
cacagtgate	gatggtcaga	taatgggagt	ttaatgacca	gtactgagag	agattatttc	41220
catggctgt	atgggccaag	aaagctgggt	ggtcagaaaag	ggacctttc	cagactctcc	41280
tgggttgt	tatttcttgc	acattgggtt	ccttcatga	ggcgccaaagg	tgtatttgt	41340
agttgtccag	tgttgcacag	ctaatggggg	gcaaaacaag	aaaggaggaa	aaaaaaagta	41400
tcctcactcc	tagttcagtg	tgttcttca	tgttggaaagg	tactgctgaa	cacagtata	41460
tgttcagtt	acaatgtaca	atgttcttgc	acattccgtc	atttgaagga	tgggtcctat	41520
ggatagtatc	atctgcatgg	tttggaaaac	aaaagatata	agctaatttt	gccctgtcta	41580
gtgactacga	gacagggaga	gaaaatctga	atatttggta	aagtagacac	agacccataa	41640
attgaaaagg	acactaatcc	tgccttagga	gacagttaagg	cacttgccc	tgctatctat	41700
tagctgtgt	tccttggaca	gctcattgtc	tctttcttag	ccaccgttac	acgttatctta	41760
tcaagtgagc	aggtgtaca	ctagatgaat	tcacagggtcc	tttttccaaag	tgcctttctta	41820
atcttcataa	tttagataat	ctctcagtag	caaaacagttt	tacaatatgg	tcaatctgag	41880
attttttaggg	ggagaatttt	agggacatc	agaaatggca	gtagttaaaa	ggaaatagga	41940
ctcacaggt	gacttcttc	tataactca	catggtagaa	gggatgagag	tttctctgg	42000
agccttttgt	attaggtc	taatccaaa	ggccccacct	cctaagacca	tcacttttgg	42060
gattggggtt	ttaaacatgt	tttagaggg	cataaaacctt	catctcattt	cattagagat	42120
atttgaaagc	tcagctcag	tgttattctc	cacagctcac	atgtatttct	ccatcacccca	42180
acctgatg	tttgaacgtt	gtagacataa	atctttcat	cattatcaag	aatattgcca	42240
aaagcttctc	agaaaattatg	aggggttttt	ttagttctta	aaatatttctc	aaagaaaagtc	42300
ccatgtacta	atgttgcct	tttgatgaaa	aaggatgaaa	tcttaatgtat	tgccttaata	42360
agctcaacaa	tgctgttag	tttagtcttc	ttatttgc	gattcttata	aacaacaaca	42420
ttcagttataa	acattaatgt	atgtgattca	ctaagttttt	tgcatgattt	tctgtgaggg	42480
ttcagatgtc	tcttggatta	tgtgtctttt	tttcatcgcc	agcatctgt	ggcagctgt	42540
actcatttgc	ttctatgggt	atgctagaag	cgagaatccg	tataactacc	aacaatttctc	42600
agaccccaat	cctaagccct	caggagggtt	tgttcttgc	ccagttgtct	caaggttaagt	42660
gttgcatttc	agacaccatt	tatgagctat	ttacctgtgt	gcagctggct	gttggggca	42720
aaggccaaag	gatgtatgcag	tagagaggg	gcagttgtca	tagtcagaaa	atctgagttc	42780
aagtctggcc	ctatca	ttaatggat	attgctcatg	gaatttactt	taccatccag	42840
caaaaatgtc	ataggacta	tatattggat	gagctctgt	tgttacat	attgcctaa	42900
aatgtctcaa	actctgagaa	gtagtaagta	taatccctat	ttatggccgg	ggaacaggaa	42960
ctaagaaaatt	tttctaataa	tttgaaggtc	tcacagctt	tagcatttgg	gtttcacttc	43020
taatcatctg	ctccaaaacc	caacttttat	taaaaactata	ctaacaactgg	tttctctctg	43080
ggagaatttt	aaaattctgt	acttagggct	gggcacagt	gcttattgcct	ataaccctat	43140
cactttggga	ggctgagatg	ggtgaatctc	ttgagtccta	gagtttgaga	ccagcctgg	43200
caacacggcg	aaaccccttc	tctattaaaa	atacaaaaaaa	ttagctgggc	gtgggtgtgt	43260
gtgcttgtag	tcccagctat	tcaggaggcg	gaggtgttaag	aatcacctga	gcccgaggagg	43320
tcaaggctgc	agtgagccga	tatcatgcca	ctgcactcca	gcctgggcaa	acggagttag	43380
gccctgttat	gaaaaaaaaa	aaatctgtac	ttaggcttcc	agatcagggt	gtatgtgtat	43440
tatgtcgaaa	acacagctat	aattgattga	gggagaaaacg	ttaccatttt	aaaggttatg	43500
ctttcaaggcc	cagatttggc	cacttagaat	ttcccgatc	acttagtaaa	ctgctgtat	43560
gtgattttt	gccagtggc	ctttcatttc	ttctaaaata	tgtactacta	gttgtgactt	43620
gtaggctata	ggggctataa	tatacaaga	caatctttat	cctcatgaag	tttacaggtt	43680
agtaagagat	agagattaa	taattataac	aacagagtga	agaaacagtga	agaaaaaaagta	43740
cagagttata	atatatataa	tagggccagg	actgcatgag	gaaggtagga	aagacatttc	43800
ggcaagagg	tgtcaggaa	aagacttgc	tgagaaaagag	ccaagttgt	gggtctggct	43860
gcttagcaat	gaccataata	cctaactttt	gcttattttt	catgaagtaa	ctaatttaac	43920
cctatgagga	aagtactact	accatctaga	tttacaggat	aagtaagcag	agatacagag	43980
aagttaaact	cttcacacgg	cttggcttt	aaacctataat	aggcttcaga	gcctccccac	44040
ttaaccactt	tgccatagct	acatccat	taggtgttac	gtagatatact	gttaagtata	44100
aggaggatga	aaggatagtt	agcttagttgg	aaaatggat	gatgaatgaa	gtgtatgtt	44160
agctaagaac	aactttcagg	ggtaacatgc	aaagaataat	ggagccaaaga	agaaaaaaata	44220
gaaaatggga	taatcccttt	tctactaagg	ggtaaccat	tgtgttattt	atcttcaggc	44280
tgtgaaggcg	gcttccata	ccttatttgca	ggaaaggatc	cccaagattt	ttggctgg	44340
gaagaaggct	gcttccata	cacaggca	gattctccat	gcaaaaatgaa	ggaagactgc	44400
tttcgttatt	actctctgt	gttaccat	gttaggaggtt	tctatggagg	ctgcaatgaa	44460
gccctgtat	agcttgagtt	ggtccatcat	ggccccatgg	cagttgtttt	tgaagtata	44520
gatgacttcc	tccactacaa	aaaggggatc	taccaccaca	ctggcttaag	agacccttc	44580
aaccctttg	agctgactaa	tcatgctgtt	ctgcttgg	gctatggcac	tgactcagcc	44640

tctggatgg	attactggat	tgttaaaaac	agctggggca	ccggctgggg	tgagaatggc	44700
tacttccgga	tccgcagagg	aactgatgag	tgtcaattt	agagcatagc	agtggcagcc	44760
acaccaattc	ctaaatttta	gggtatgcct	tccagtattt	cataatgatc	tgcatacggt	44820
gtaaagggga	attggatat	tcacagactg	tagactttca	gcagcaatct	cagaagctta	44880
caaatacgatt	tccatgaaga	tatttgtctt	cagaattaaa	actgcctta	attttaatat	44940
accttcaat	cggccactgg	ccatTTTTT	ctaagtattt	aattaatgg	gaattttctg	45000
gaagatggc	agctatgaag	taatagagtt	tgcttaatca	tttgtaattt	aaacatgcta	45060
tatTTTTAA	aatcaatgtg	aaaacataga	cttatttttta	aattgtacca	atcacaagaa	45120
aataatggca	ataattatca	aaacttttaa	aatagatgct	catattttta	aaataaaagtt	45180
ttaaaa						45186

<210> 2
<211> 1838
<212> DNA
<213> Homo sapiens

<400> 2						
aattcttcac	ctcttttctc	agctccctgc	agcatgggtg	ctggggccctc	cttgctgctc	60
gcccgcctcc	tgctgcttct	ctccggcgac	ggcgccgtgc	gctgcgacac	acctgccaac	120
tgcaccttate	ttgacctgtct	gggcacctgg	gttccagg	ttgggtccag	cggttccag	180
cgcgatgtca	actgctcggt	tatggacca	caagaaaaaa	aagttagtgg	gtaccttcag	240
aagctggata	cagcatatga	tgaccttggc	aattctggcc	atttaccat	catttacaac	300
caaggcttg	agattgtgtt	gaatgactac	aagtggttt	cctttttaa	gtataaagaa	360
gagggcagca	aggtgaccac	ttactgcaac	gagacaatga	ctgggtgggt	gcatgatgt	420
ttggggccgga	actggggcttg	tttcacccga	aagaagggtgg	gaactgcctc	tgagaatgt	480
tatgtcaaca	cagcacacact	taagaattct	caggaaaagt	attctaatag	gctctacaag	540
tatgatcaca	acttgtgaa	agctatcaat	gccattcaga	agtcttggac	tgcaactaca	600
tacatggaaat	atgagactct	tacccttggga	gatatgatta	ggagaagtgg	tggccacagt	660
cgaaaaatcc	caaggcccaa	acctgcacca	ctgactgctg	aaatacagca	aaagattttg	720
catttgcCAA	catcttggga	ctggagaaat	gttcatggta	tcaatTTTGT	cagtcctgtt	780
cgaaaaccaag	cattctgtgg	cagctgtac	tcatTTGTT	ctatgggtat	gctagaagcg	840
agaatccgtA	tactaaccAA	caatttcag	accccaatcc	taagccctca	ggaggTTG	900
tctttagcc	agtagtgcA	aggctgtgaa	ggcggcttcc	cataccttat	tgccaggaaag	960
tacggccaaag	atttgggct	ggtgaaagaa	gcttgcttcc	cctacacagg	cactgattt	1020
ccatgcAAA	tgaaggaaga	ctgcttcgt	tattactct	ctgagtagca	ctatgttagga	1080
ggtttctatg	gaggctgcaa	tgaagccctg	atgaagctt	agttggtcca	tcatggggcc	1140
atggcagttg	cttttgaagt	atatgtac	ttcttccact	acaaaaagg	gatctaccac	1200
cacactggc	taagagaccc	tttcaacccc	tttgagctga	ctaattcatgc	tgttctgtt	1260
gtgggctatg	gcactgactc	agcctctgg	atggattact	ggattgttAA	aaacagctgg	1320
ggcaccggct	ggggtgagaa	tggctacttc	cggatccgca	gaggaactga	tgagtgtca	1380
attgagagca	tagcagtggc	agccacacca	attcctaaat	tgttagggtat	gccttccagt	1440
atttcataat	gatctgcac	agttgtaaag	ggaaatgg	atattcacag	actgttagact	1500
ttcagcagca	atctcagaag	cttacaaata	gatttccat	aaagatattt	tcttcagaat	1560
taaaaactgCC	cttaatttttA	ataacctt	caatcggcca	ctggccattt	ttttctaagt	1620
attcaattaa	gtggaaat	tctggaaat	ggtcagctat	gaagaaatag	agtttgctt	1680
atcattttgtA	attcaaacat	gtctatTTT	ttaaaatcaa	tgtgaaaaca	tagacttatt	1740
tttaaaatgt	accaatcaca	agaaaataat	ggcaataatt	atcaaaactt	ttaaaataga	1800
tgctcatatt	tttaaaataaa	agttttaaaa	ataactgc			1838

<210> 3
<211> 463
<212> PRT
<213> Homo sapiens

<400> 3															
Met	Gly	Ala	Gly	Pro	Ser	Leu	Leu	Leu	Leu	Leu	Leu				
1				5			10			15					
Ser	Gly	Asp	Gly	Ala	Val	Arg	Cys	Asp	Thr	Pro	Ala	Asn	Cys	Thr	Tyr
					20			25			30				
Leu	Asp	Leu	Leu	Gly	Thr	Trp	Val	Phe	Gln	Val	Gly	Ser	Ser	Gly	Ser
					35			40			45				
Gln	Arg	Asp	Val	Asn	Cys	Ser	Val	Met	Gly	Pro	Gln	Glu	Lys	Lys	Val
							55			60					

Val	Val	Tyr	Leu	Gln	Lys	Leu	Asp	Thr	Ala	Tyr	Asp	Asp	Leu	Gly	Asn
65						70				75					80
Ser	Gly	His	Phe	Thr	Ile	Ile	Tyr	Asn	Gln	Gly	Phe	Glu	Ile	Val	Leu
						85				90					95
Asn	Asp	Tyr	Lys	Trp	Phe	Ala	Phe	Phe	Lys	Tyr	Lys	Glu	Glu	Gly	Ser
						100				105					110
Lys	Val	Thr	Thr	Tyr	Cys	Asn	Glu	Thr	Met	Thr	Gly	Trp	Val	His	Asp
						115				120					125
Val	Leu	Gly	Arg	Asn	Trp	Ala	Cys	Phe	Thr	Gly	Lys	Lys	Val	Gly	Thr
						130				135					140
Ala	Ser	Glu	Asn	Val	Tyr	Val	Asn	Thr	Ala	His	Leu	Lys	Asn	Ser	Gln
						145				150					160
Glu	Lys	Tyr	Ser	Asn	Arg	Leu	Tyr	Lys	Tyr	Asp	His	Asn	Phe	Val	Lys
						165				170					175
Ala	Ile	Asn	Ala	Ile	Gln	Lys	Ser	Trp	Thr	Ala	Thr	Thr	Tyr	Met	Glu
						180				185					190
Tyr	Glu	Thr	Leu	Thr	Leu	Gly	Asp	Met	Ile	Arg	Arg	Ser	Gly	Gly	His
						195				200					205
Ser	Arg	Lys	Ile	Pro	Arg	Pro	Lys	Pro	Ala	Pro	Leu	Thr	Ala	Glu	Ile
						210				215					220
Gln	Gln	Lys	Ile	Leu	His	Leu	Pro	Thr	Ser	Trp	Asp	Trp	Arg	Asn	Val
						225				230					240
His	Gly	Ile	Asn	Phe	Val	Ser	Pro	Val	Arg	Asn	Gln	Ala	Ser	Cys	Gly
						245				250					255
Ser	Cys	Tyr	Ser	Phe	Ala	Ser	Met	Gly	Met	Leu	Glu	Ala	Arg	Ile	Arg
						260				265					270
Ile	Leu	Thr	Asn	Asn	Ser	Gln	Thr	Pro	Ile	Leu	Ser	Pro	Gln	Glu	Val
						275				280					285
Val	Ser	Cys	Ser	Gln	Tyr	Ala	Gln	Gly	Cys	Glu	Gly	Gly	Phe	Pro	Tyr
						290				295					300
Leu	Ile	Ala	Gly	Lys	Tyr	Ala	Gln	Asp	Phe	Gly	Leu	Val	Glu	Glu	Ala
						305				310					320
Cys	Phe	Pro	Tyr	Thr	Gly	Thr	Asp	Ser	Pro	Cys	Lys	Met	Lys	Glu	Asp
						325				330					335
Cys	Phe	Arg	Tyr	Tyr	Ser	Ser	Glu	Tyr	His	Tyr	Val	Gly	Gly	Phe	Tyr
						340				345					350
Gly	Gly	Cys	Asn	Glu	Ala	Leu	Met	Lys	Leu	Glu	Leu	Val	His	His	Gly
						355				360					365
Pro	Met	Ala	Val	Ala	Phe	Glu	Val	Tyr	Asp	Asp	Phe	Leu	His	Tyr	Lys
						370				375					380
Lys	Gly	Ile	Tyr	His	His	Thr	Gly	Leu	Arg	Asp	Pro	Phe	Asn	Pro	Phe
						385				390					400
Glu	Leu	Thr	Asn	His	Ala	Val	Leu	Leu	Val	Gly	Tyr	Gly	Thr	Asp	Ser
						405				410					415
Ala	Ser	Gly	Met	Asp	Tyr	Trp	Ile	Val	Lys	Asn	Ser	Trp	Gly	Thr	Gly
						420				425					430
Trp	Gly	Glu	Asn	Gly	Tyr	Phe	Arg	Ile	Arg	Arg	Gly	Thr	Asp	Glu	Cys
						435				440					445
Ala	Ile	Glu	Ser	Ile	Ala	Val	Ala	Ala	Thr	Pro	Ile	Pro	Lys	Leu	
						450				455					460

<210> 4
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Primer

<400> 4
aggagttgt gtctttagc c

```

<210> 5
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Primer

<400> 5
agtgcctgtg taggggaaggc

<210> 6
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Primer

<400> 6
tgtcaactgc tcggatcgatgg gttaa

<210> 7
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Primer

<400> 7
tcgagcttct cttcgatcacac cact

<210> 8
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Primer

<400> 8
tgactacaag tggtttgcct tttt

<210> 9
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Primer

<400> 9
tgctgccctc ttctttataac tgc

<210> 10
<211> 24
<212> DNA
<213> Artificial Sequence

```

<220>
<223> Description of Artificial Sequence: Primer

<400> 10
gcctctgaga atgtgtatgt caac

<210> 11
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Primer

<400> 11
cctgccccaa aaatgagata

<210> 12
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Primer

<400> 12
tcgaaaaatc ccaaggtaat c

<210> 13
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Primer

<400> 13
gggcctagaa aggaaaatata catt

<210> 14
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Primer

<400> 14
aatttgttcg gaactattta ttga

<210> 15
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Primer

24

20

21

24

24

<pre> <400> 15 tcgcttctag cataccccata </pre>	20
<pre> <210> 16 <211> 24 <212> DNA <213> Artificial Sequence <220> <223> Description of Artificial Sequence: Primer </pre>	
<pre> <400> 16 ggcgatcaga ctggcacact agaa </pre>	24
<pre> <210> 17 <211> 24 <212> DNA <213> Artificial Sequence <220> <223> Description of Artificial Sequence: Primer </pre>	
<pre> <400> 17 cttaccacata accgagcagt tgac </pre>	24
<pre> <210> 18 <211> 24 <212> DNA <213> Artificial Sequence <220> <223> Description of Artificial Sequence: Primer </pre>	
<pre> <400> 18 gcagactgtg ctc当地actgg gtag </pre>	24
<pre> <210> 19 <211> 24 <212> DNA <213> Artificial Sequence <220> <223> Description of Artificial Sequence: Primer </pre>	
<pre> <400> 19 tctactaatac agaagagggtt tcag </pre>	24
<pre> <210> 20 <211> 23 <212> DNA <213> Artificial Sequence <220> <223> Description of Artificial Sequence: Primer </pre>	
<pre> <400> 20 ggcacattta ctgtgaatga gag </pre>	23

```

<210> 21
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Primer

<400> 21
gtctcatttg tagcaactca c 21

<210> 22
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Primer

<400> 22
ccactttcca cttaggcaca g 21

<210> 23
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Primer

<400> 23
aggatggtat tcagcattca ta 22

<210> 24
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Primer

<400> 24
atcctagcta gtctggtagc tgaa 24

<210> 25
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Primer

<400> 25
tctaggatac cccgaaatcc atca 24

<210> 26
<211> 24
<212> DNA
<213> Artificial Sequence

```

<220>
<223> Description of Artificial Sequence: Primer

<400> 26
gattctctgt gaggcttcag atgt 24

<210> 27
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Primer

<400> 27
gccaaacaaca gccagctgca caca 24

<210> 28
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Primer

<400> 28
tccccactta accactttgc 20

<210> 29
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Primer

<400> 29
acttcatagc tgaccatctt cc 22

<210> 30
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Primer

<400> 30
ggcgccctcc tgctgcttct 20

<210> 31
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Primer

<400> 31
ggcttaggt tgggggtctga 20

gtatgctaga agcgagaatc cgtat 25

ccaatgctaa aacttggtaa gacc 24